Journal of Privacy and Confidentiality
Vol. 9 (1) 2019

TPDP 2016

Submitted March 2018
Published March 2019

APPENDIX TO: DIFFERENTIALLY PRIVATE CONFIDENCE INTERVALS FOR EMPIRICAL RISK MINIMIZATION

YUE WANG, DANIEL KIFER, AND JAEWOO LEE

The main manuscript is available at https://doi.org/10.29012/jpc.660.

Appendix B. Complete Experimental Results

B.1. **Allocation for the Privacy Budget.** See Figures 1 to 12.

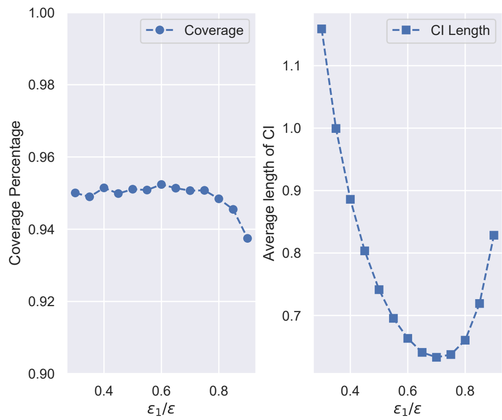
B.2. **Empirical Sample Complexity of Private Confidence Intervals.** See Figures 13 to 20.

Key words and phrases: Differential Privacy, Objective Perturbation, Output Perturbation, Confidence Intervals.

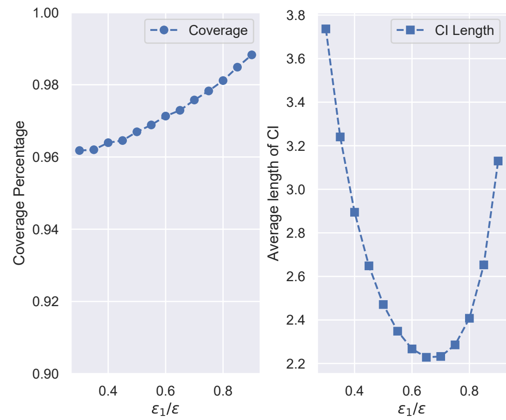
JPrivacy Confidentiality

www.journalprivacyconfidentiality.org
DOI:10.29012/jpc.660

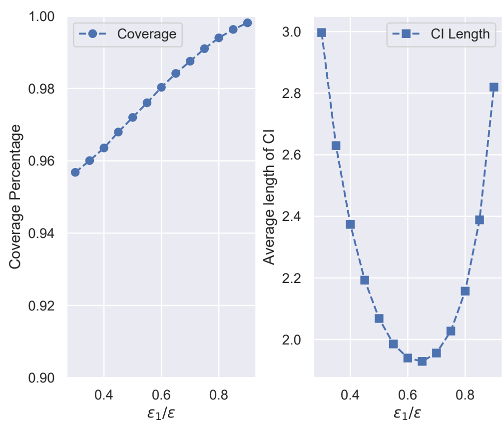
© Y. Wang, D. Kifer, and J. Lee
Creative Commons (CC BY-NC-ND 4.0)



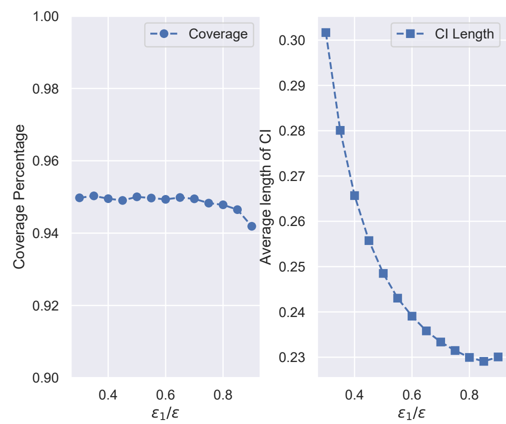
(A) Adult, $n = 5,000$, $d = 1$



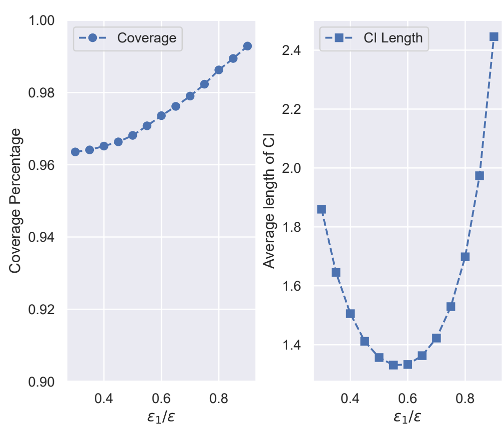
(B) Adult, $n = 5,000$, $d = 5$



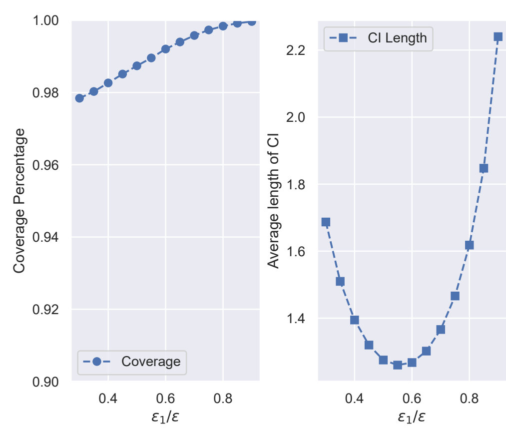
(C) Adult, $n = 10,000$, $d = 10$



(D) US, $n = 5,000$, $d = 1$

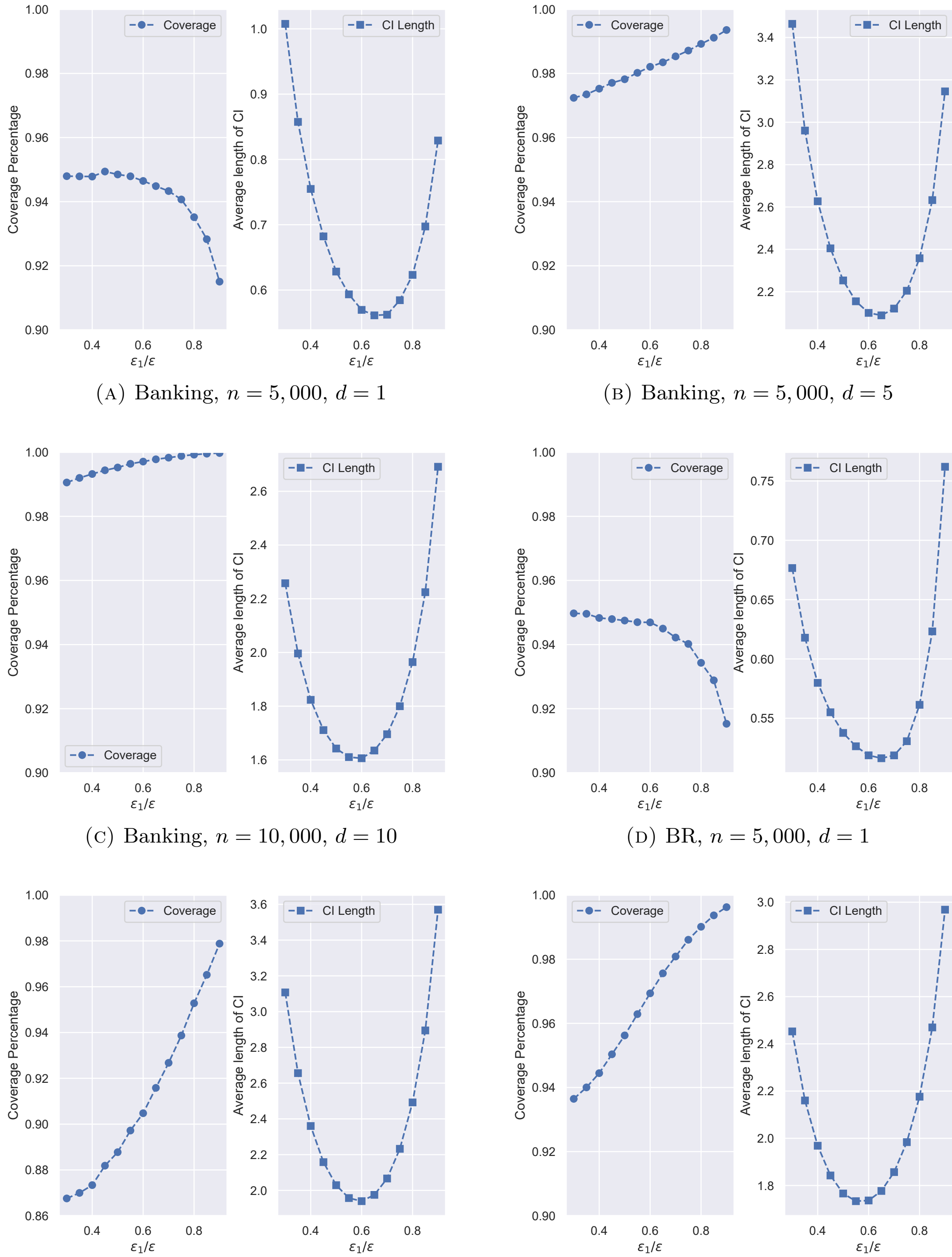


(E) US, $n = 5,000$, $d = 5$



(F) US, $n = 10,000$, $d = 10$

Figure 1: **[ϵ-DP, objective perturbation, logistic regression]** Coverage percentage and average length of confidence intervals vs. ϵ_1/ϵ for objective perturbation based ϵ-DP confidence intervals for linear regression with a total privacy budget of $\epsilon = 1.0$. $\epsilon_2 = \epsilon_3 = (\epsilon - \epsilon_1)/2$, $c = 0.001$.



(A) Banking, $n = 5,000$, $d = 1$

(B) Banking, $n = 5,000$, $d = 5$

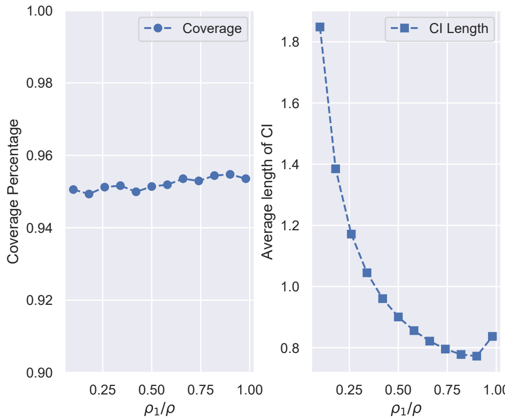
(C) Banking, $n = 10,000$, $d = 10$

(D) BR, $n = 5,000$, $d = 1$

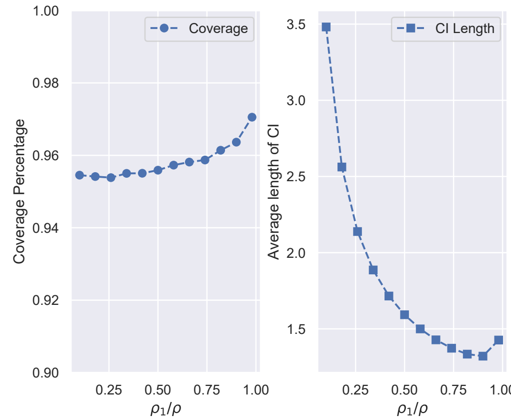
(E) BR, $n = 5,000$, $d = 5$

(F) BR, $n = 10,000$, $d = 10$

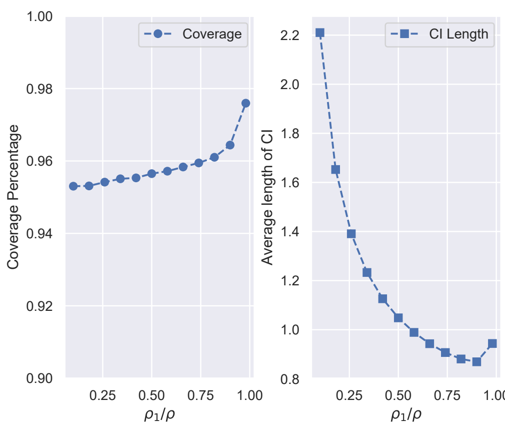
Figure 2: **[ϵ-DP, objective perturbation, SVM]** Coverage percentage and average length of confidence intervals vs. ϵ_1/ϵ for objective perturbation based ϵ-DP confidence intervals for SVM with a total privacy budget of $\epsilon = 1.0$. $\epsilon_2 = \epsilon_3 = (\epsilon - \epsilon_1)/2$, $c = 0.001$, $h = 1.0$.



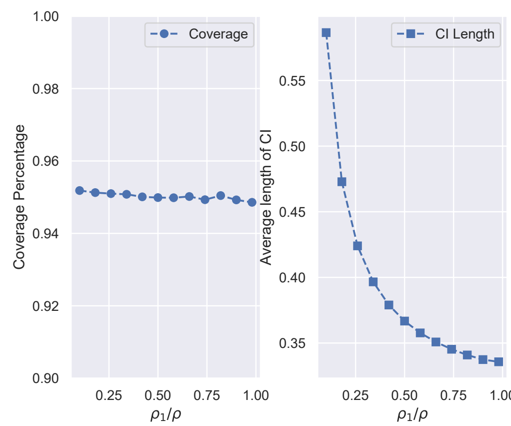
(A) Banking, $n = 5,000$, $d = 1$



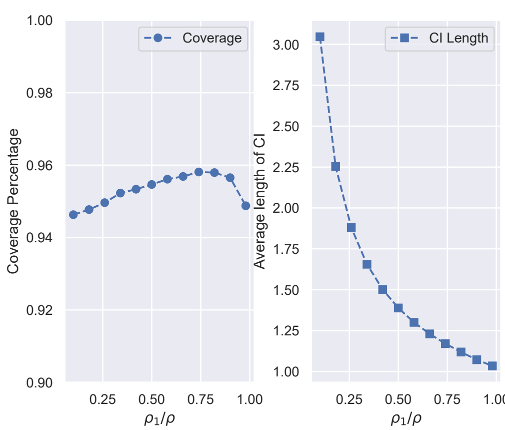
(B) Banking, $n = 5,000$, $d = 5$



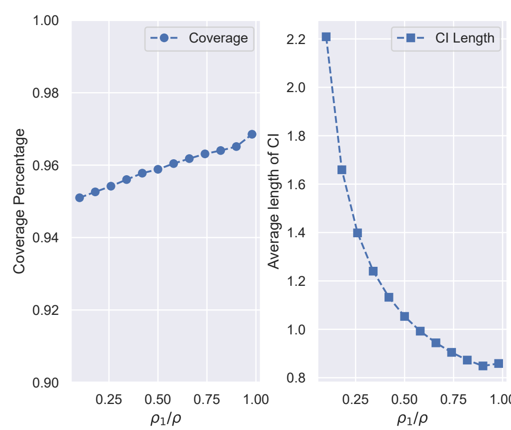
(C) Banking, $n = 10,000$, $d = 10$



(D) BR, $n = 5,000$, $d = 1$

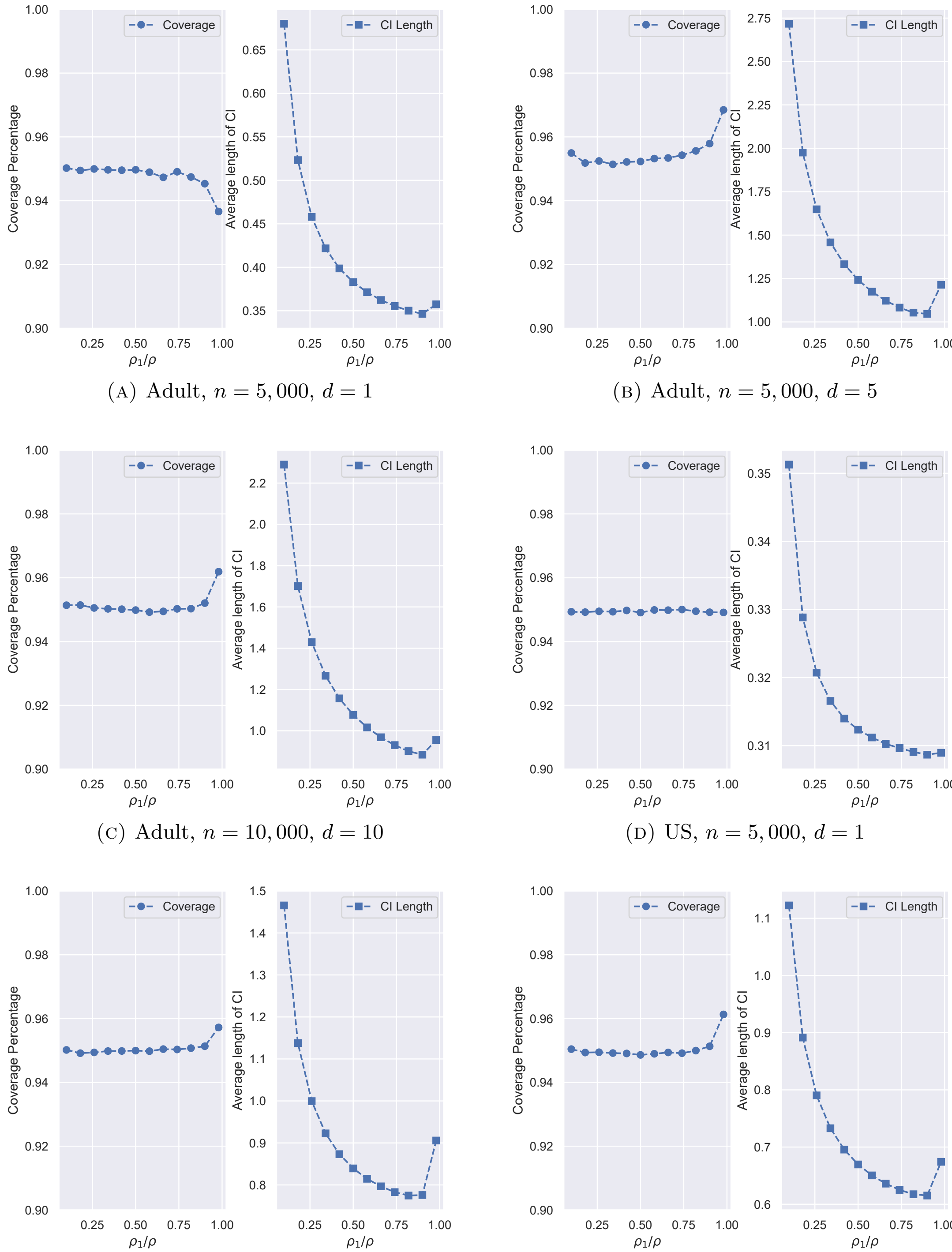


(E) BR, $n = 5,000$, $d = 5$



(F) BR, $n = 10,000$, $d = 10$

Figure 3: **[zCDP, objective perturbation, logistic regression]** Coverage percentage and average length of confidence intervals vs. ρ_1/ρ for objective perturbation based zCDP confidence intervals for linear regression with a total privacy budget of $\rho = 0.5$. $\rho_2 = \rho_3 = (\rho - \rho_1)/2$, $c = 0.001$.



(A) Adult, $n = 5,000$, $d = 1$

(B) Adult, $n = 5,000$, $d = 5$

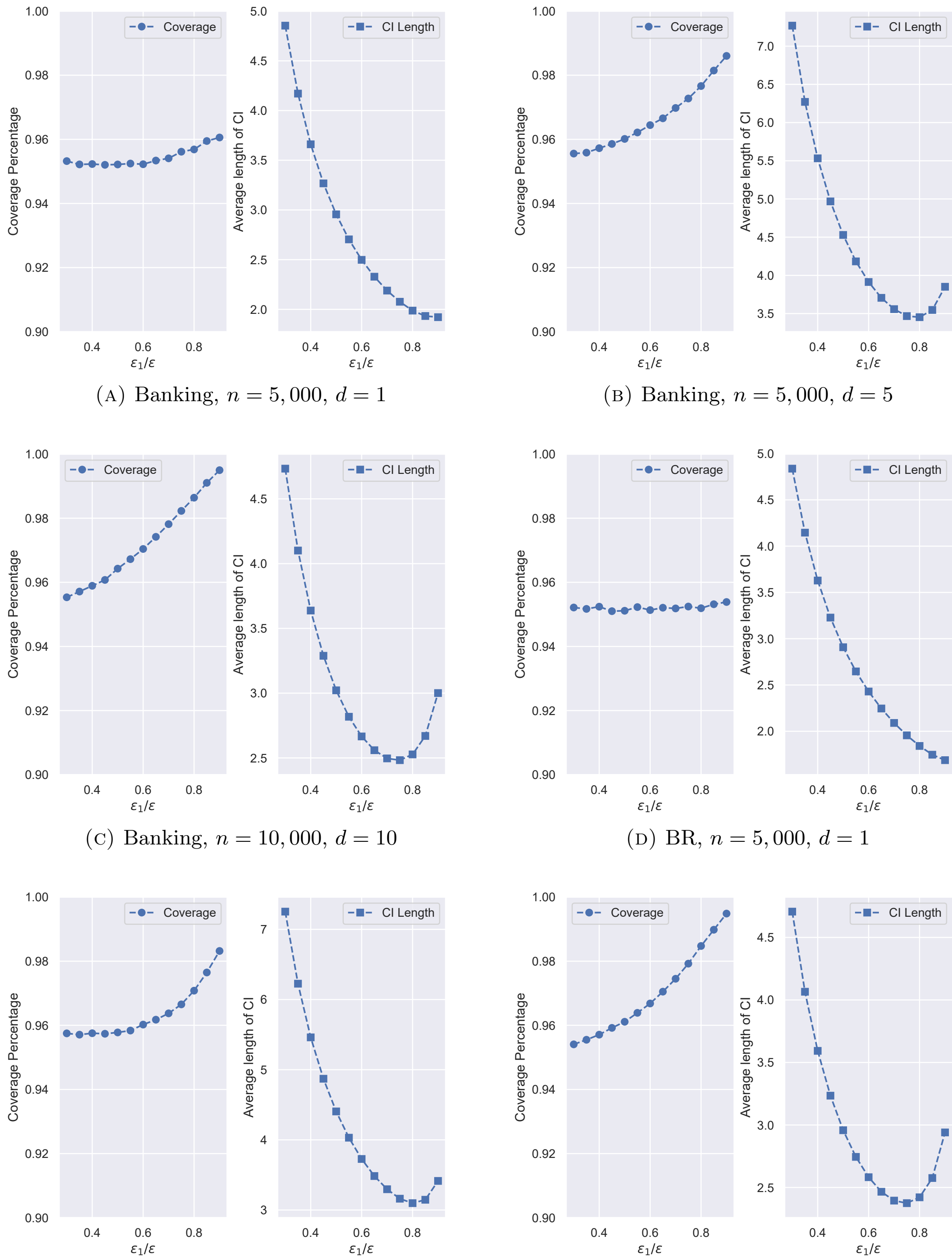
(C) Adult, $n = 10,000$, $d = 10$

(D) US, $n = 5,000$, $d = 1$

(E) US, $n = 5,000$, $d = 5$

(F) US, $n = 10,000$, $d = 10$

Figure 4: **[zCDP, objective perturbation, SVM]** Coverage percentage and average length of confidence intervals vs. ρ_1/ρ for objective perturbation based zCDP confidence intervals for SVM with a total privacy budget of $\rho = 0.5$. $\rho_2 = \rho_3 = (\rho - \rho_1)/2$, $c = 0.001$, $h = 1.0$.



(A) Banking, $n = 5,000$, $d = 1$

(B) Banking, $n = 5,000$, $d = 5$

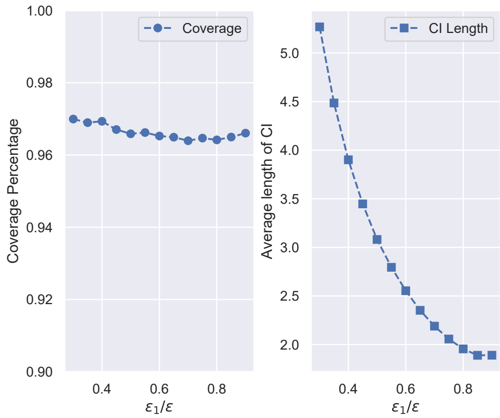
(C) Banking, $n = 10,000$, $d = 10$

(D) BR, $n = 5,000$, $d = 1$

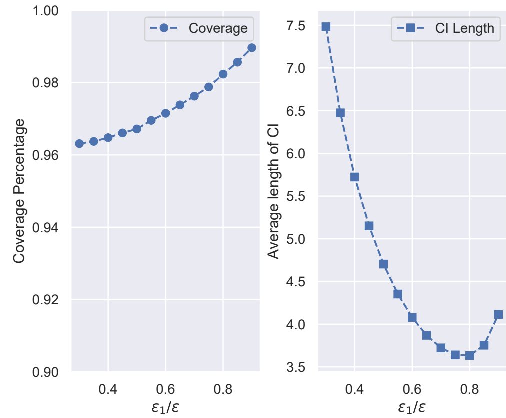
(E) BR, $n = 5,000$, $d = 5$

(F) BR, $n = 10,000$, $d = 10$

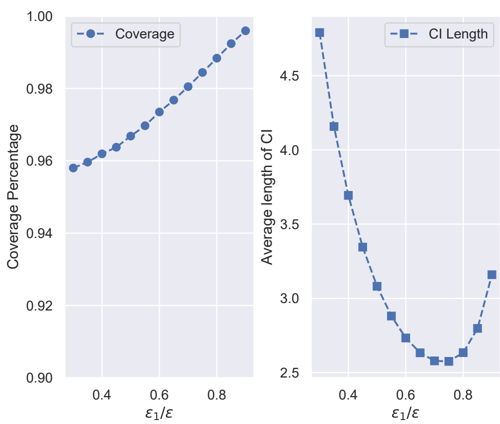
Figure 5: **[ϵ-DP, output perturbation, logistic regression]** Coverage percentage and average length of confidence intervals vs. ϵ_1/ϵ for output perturbation based ϵ-DP confidence intervals for linear regression with a total privacy budget of $\epsilon = 1.0$. $\epsilon_2 = \epsilon_3 = (\epsilon - \epsilon_1)/2$, $c = 0.001$.



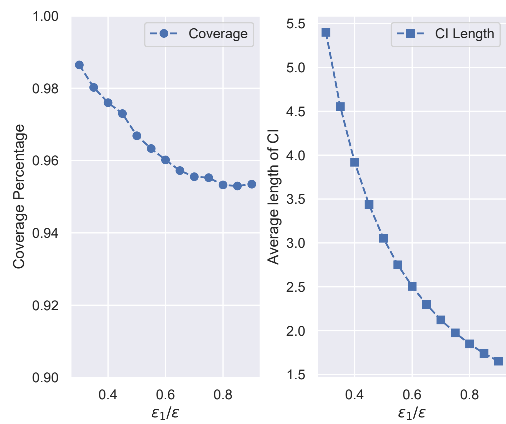
(A) Adult, $n = 5,000$, $d = 1$



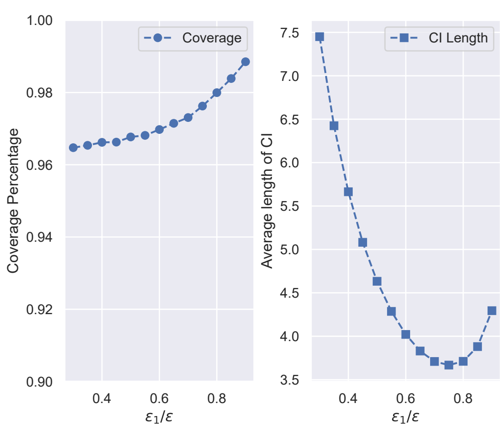
(B) Adult, $n = 5,000$, $d = 5$



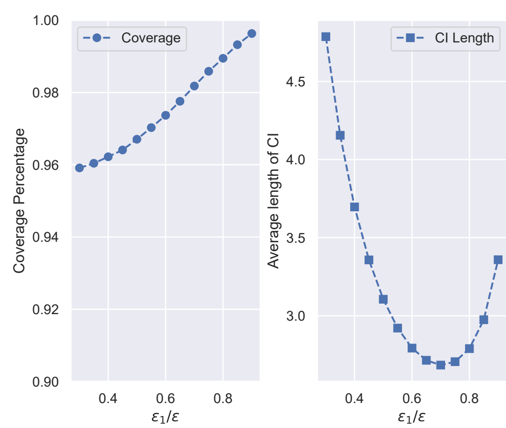
(C) Adult, $n = 10,000$, $d = 10$



(D) US, $n = 5,000$, $d = 1$

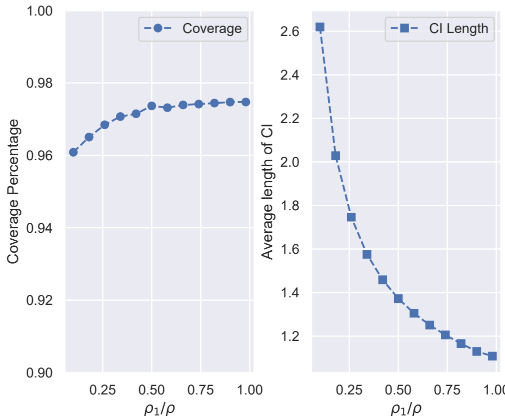


(E) US, $n = 5,000$, $d = 5$

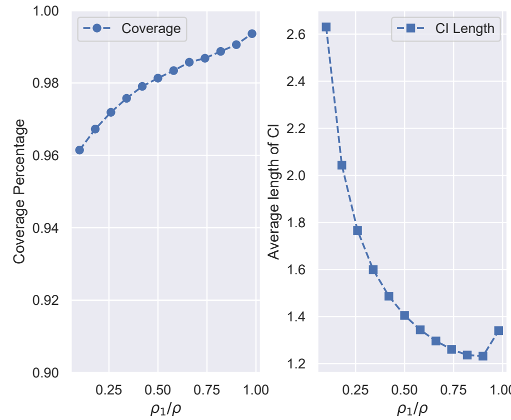


(F) US, $n = 10,000$, $d = 10$

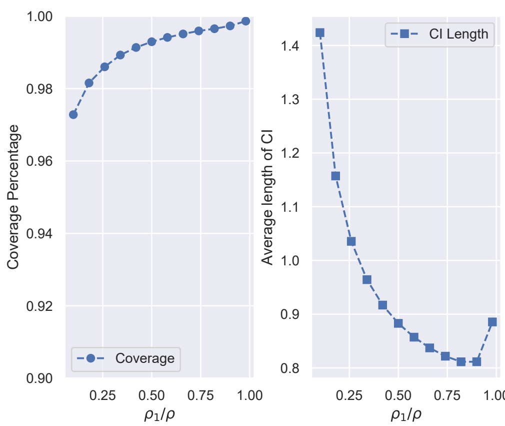
Figure 6: [**ϵ-DP, output perturbation, SVM**] Coverage percentage and average length of confidence intervals vs. ϵ_1/ϵ for output perturbation based ϵ-DP confidence intervals for SVM with a total privacy budget of $\epsilon = 1.0$. $\epsilon_2 = \epsilon_3 = (\epsilon - \epsilon_1)/2$, $c = 0.001$, $h = 1.0$.



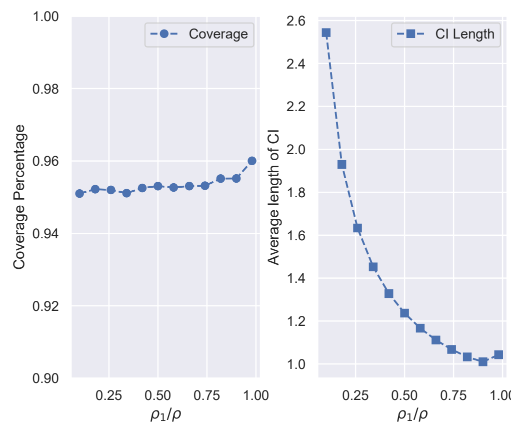
(A) KDDCUP99, $n = 5,000$, $d = 1$



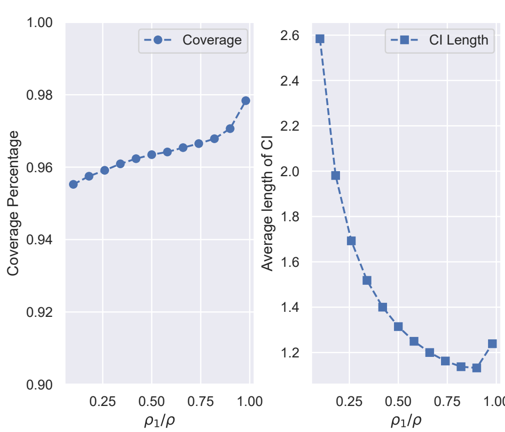
(B) KDDCUP99, $n = 5,000$, $d = 5$



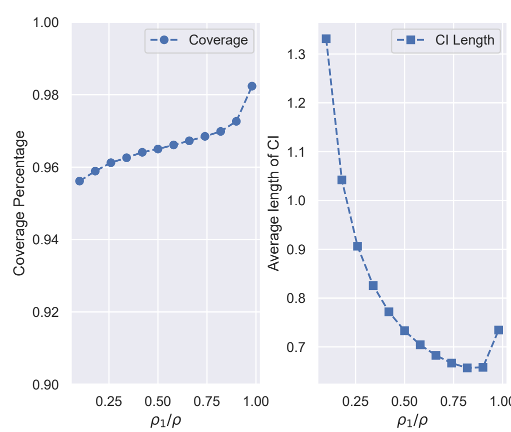
(C) KDDCUP99, $n = 10,000$, $d = 10$



(D) Banking, $n = 5,000$, $d = 1$

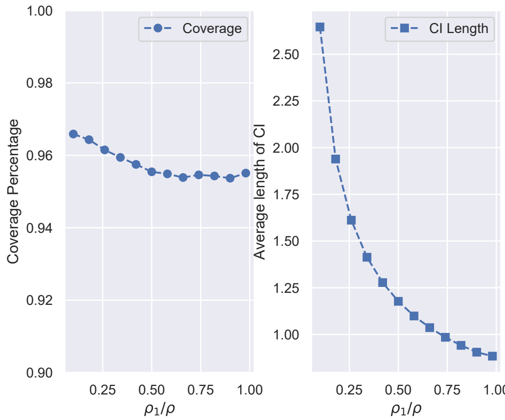


(E) Banking, $n = 5,000$, $d = 5$

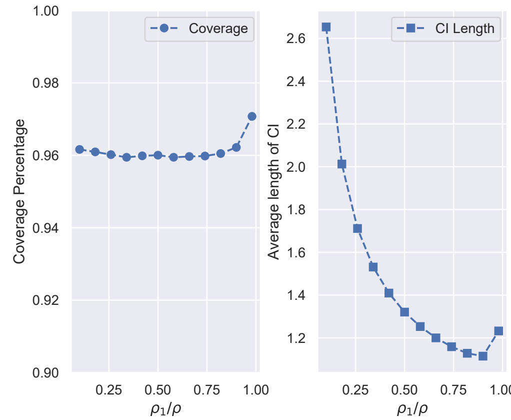


(F) Banking, $n = 10,000$, $d = 10$

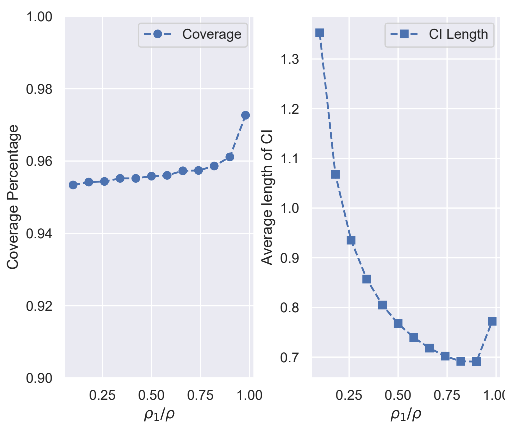
Figure 7: **[zCDP, output perturbation, logistic regression]** Coverage percentage and average length of confidence intervals vs. ρ_1/ρ for output perturbation based zCDP confidence intervals for linear regression with a total privacy budget of $\rho = 0.5$. $\rho_2 = \rho_3 = (\rho - \rho_1)/2$, $c = 0.001$.



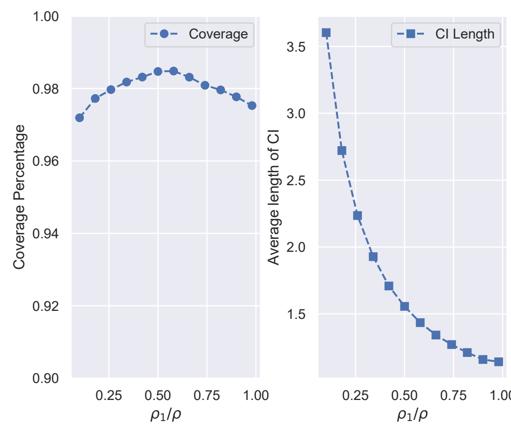
(A) Adult, $n = 5,000$, $d = 1$



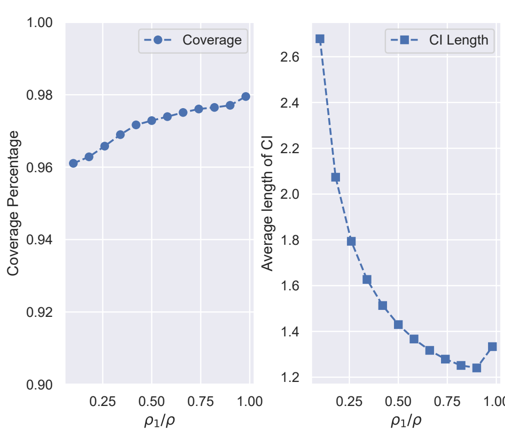
(B) Adult, $n = 5,000$, $d = 5$



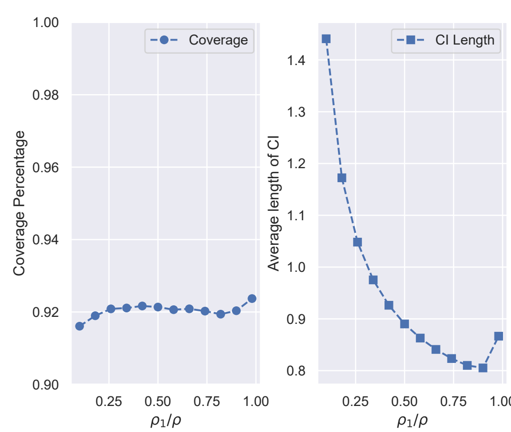
(C) Adult, $n = 10,000$, $d = 10$



(D) KDDCUP99, $n = 5,000$, $d = 1$

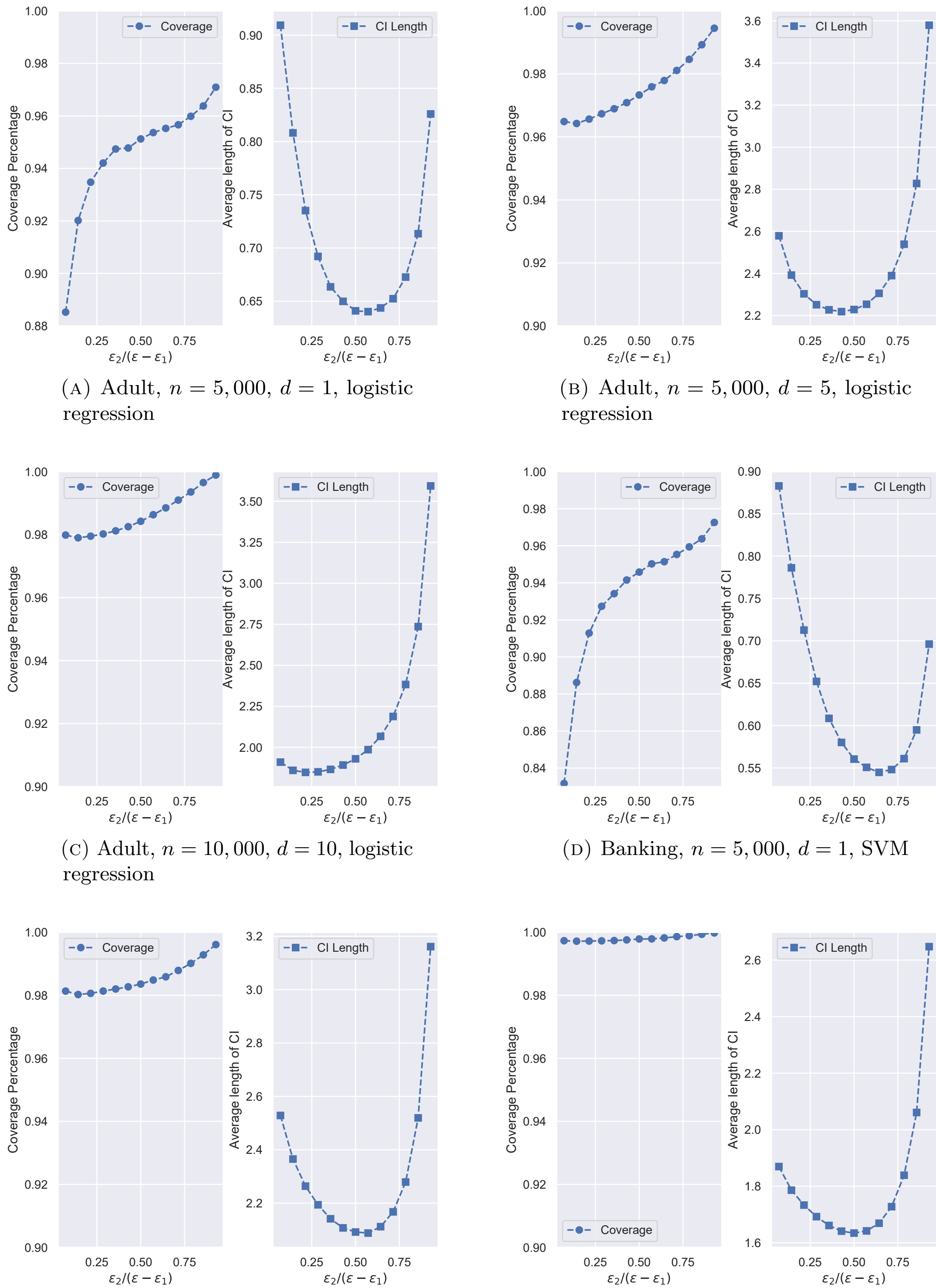


(E) KDDCUP99, $n = 5,000$, $d = 5$



(F) KDDCUP99, $n = 10,000$, $d = 10$

Figure 8: **[zCDP, output perturbation, SVM]** Coverage percentage and average length of confidence intervals vs. ρ_1/ρ for output perturbation based zCDP confidence intervals for SVM with a total privacy budget of $\rho = 0.5$. $\rho_2 = \rho_3 = (\rho - \rho_1)/2$, $c = 0.001$, $h = 1.0$.



(A) Adult, $n = 5,000$, $d = 1$, logistic regression

(B) Adult, $n = 5,000$, $d = 5$, logistic regression

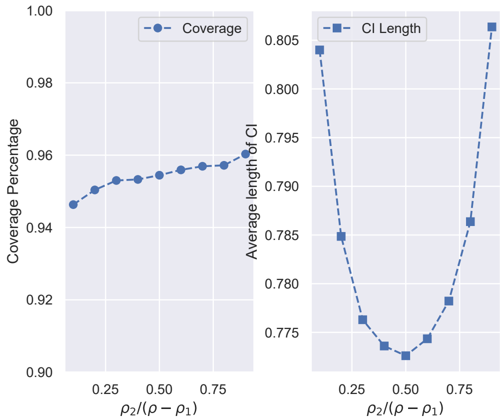
(C) Adult, $n = 10,000$, $d = 10$, logistic regression

(D) Banking, $n = 5,000$, $d = 1$, SVM

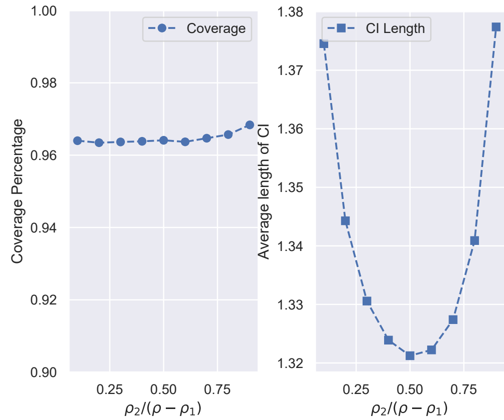
(E) Banking, $n = 5,000$, $d = 5$, SVM

(F) Banking, $n = 10,000$, $d = 10$, SVM

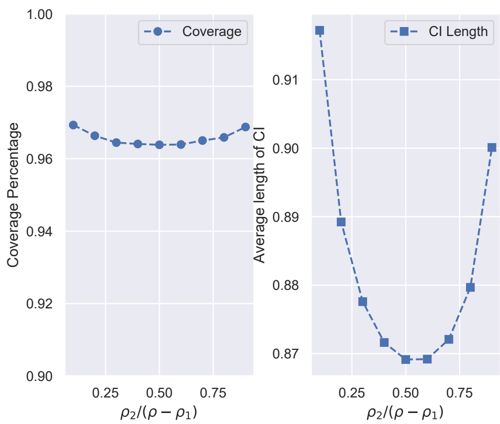
Figure 9: [**ϵ-DP, objective perturbation**] Coverage percentage and average length of confidence intervals vs. $\epsilon_2/(\epsilon - \epsilon_1)$ for objective perturbation based ϵ-DP confidence intervals with a total privacy budget of $\epsilon = 1.0$. $\epsilon_1 = 0.65$, $c = 0.001$, $h = 1.0$.



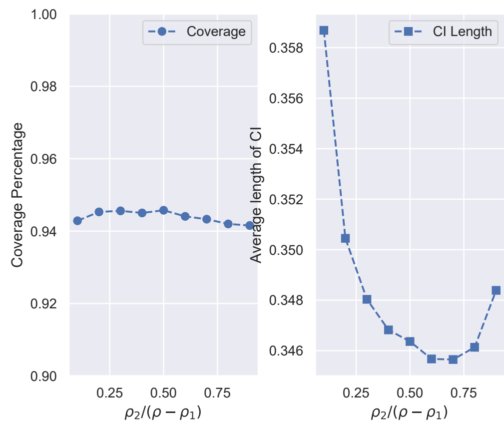
(A) Banking, $n = 5,000$, $d = 1$, logistic regression



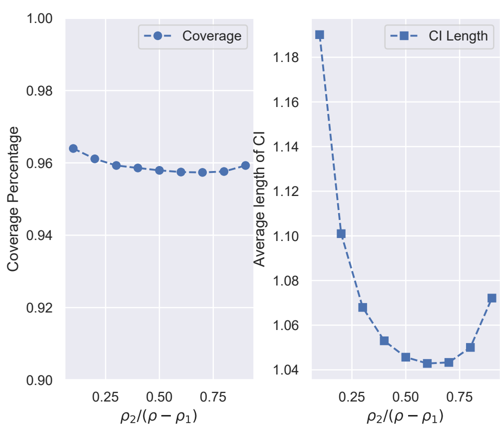
(B) Banking, $n = 5,000$, $d = 5$, logistic regression



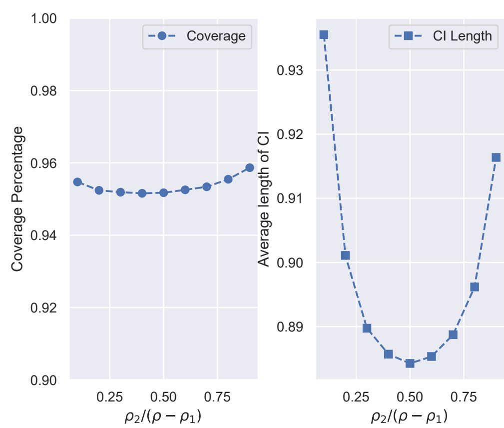
(C) Banking, $n = 10,000$, $d = 10$, logistic regression



(D) Adult, $n = 5,000$, $d = 1$, SVM

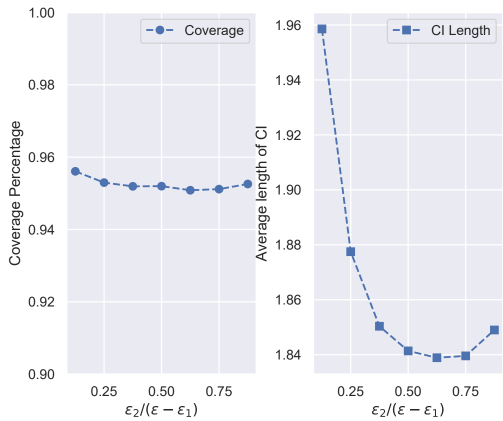


(E) Adult, $n = 5,000$, $d = 5$, SVM

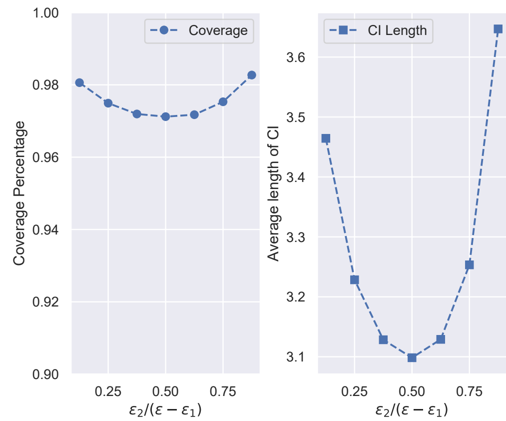


(F) Adult, $n = 10,000$, $d = 10$, SVM

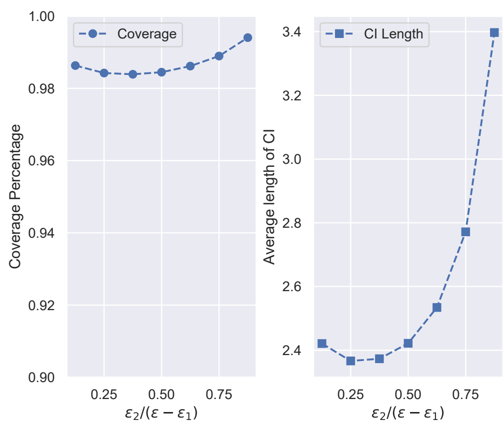
Figure 10: **[zCDP, objective perturbation]** Coverage percentage and average length of confidence intervals vs. $\rho_2/(\rho-\rho_1)$ for objective perturbation based zCDP confidence intervals with a total privacy budget of $\rho = 0.5$. $\rho_1 = 0.45$, $c = 0.001$, $h = 1.0$.



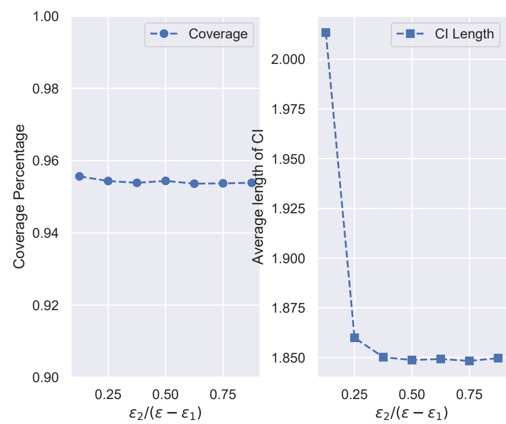
(A) BR, $n = 5,000$, $d = 1$, logistic regression



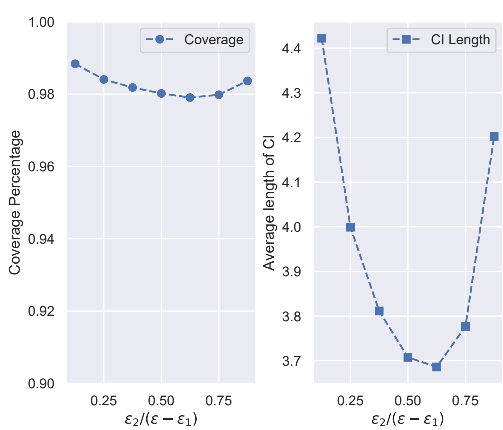
(B) BR, $n = 5,000$, $d = 5$, logistic regression



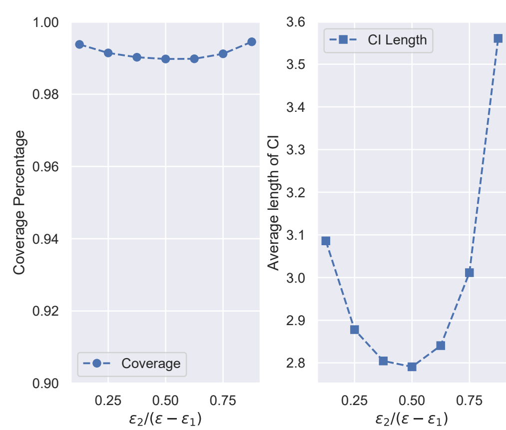
(C) BR, $n = 10,000$, $d = 10$, logistic regression



(D) US, $n = 5,000$, $d = 1$, SVM

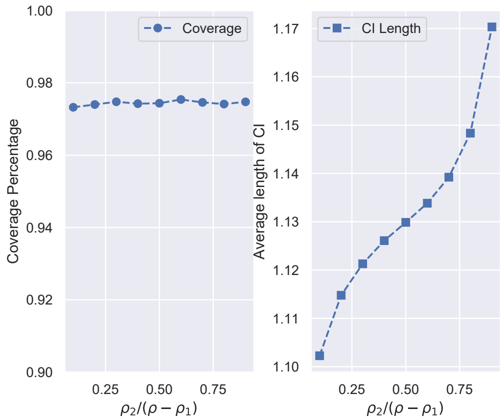


(E) US, $n = 5,000$, $d = 5$, SVM

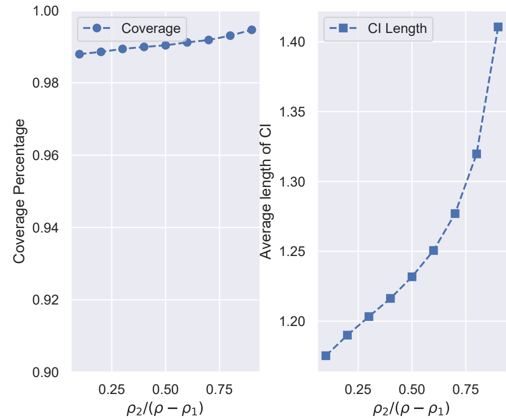


(F) US, $n = 10,000$, $d = 10$, SVM

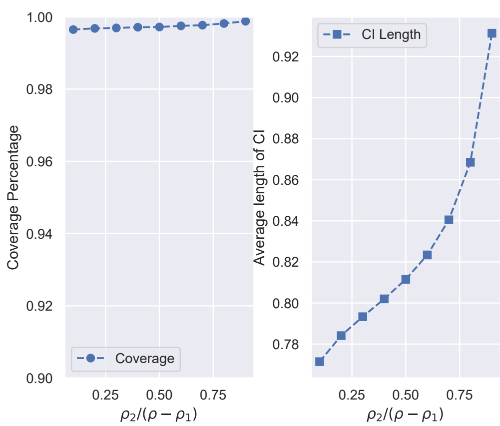
Figure 11: [**ϵ-DP, output perturbation**] Coverage percentage and average length of confidence intervals vs. $\epsilon_2/(\epsilon - \epsilon_1)$ for output perturbation based ϵ-DP confidence intervals with a total privacy budget of $\epsilon = 1.0$. $\epsilon_1 = 0.8$, $c = 0.001$, $h = 1.0$.



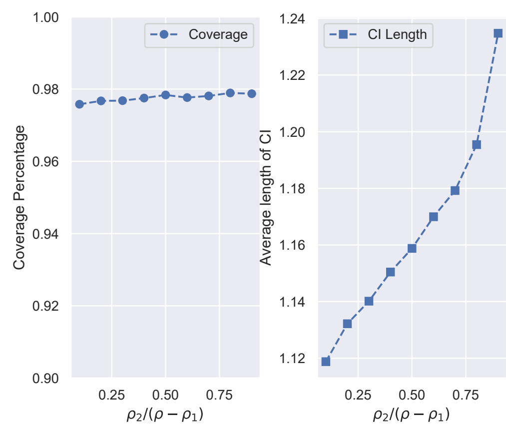
(A) KDDCUP99, $n = 5,000$, $d = 1$, logistic regression



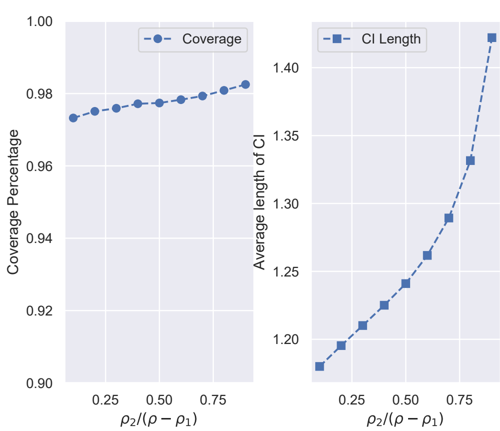
(B) KDDCUP99, $n = 5,000$, $d = 5$, logistic regression



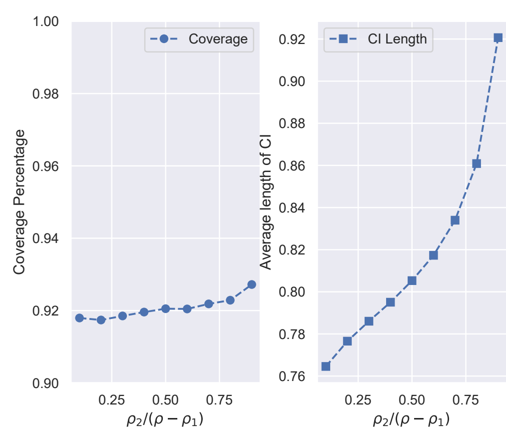
(C) KDDCUP99, $n = 10,000$, $d = 10$, logistic regression



(D) KDDCUP99, $n = 5,000$, $d = 1$, SVM

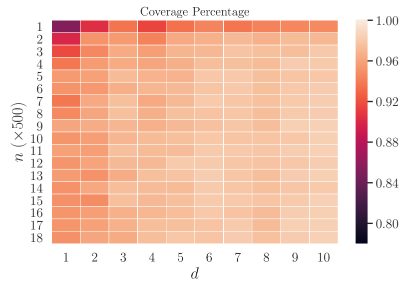


(E) KDDCUP99, $n = 5,000$, $d = 5$, SVM

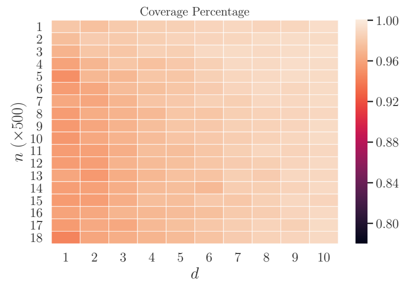


(F) KDDCUP99, $n = 10,000$, $d = 10$, SVM

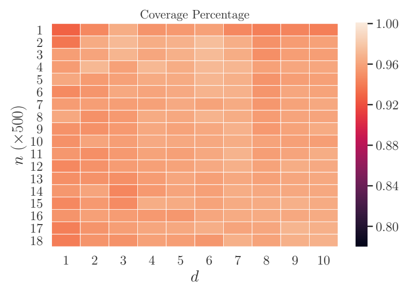
Figure 12: **[zCDP, output perturbation]** Coverage percentage and average length of confidence intervals vs. $\rho_2/(\rho - \rho_1)$ for output perturbation based zCDP confidence intervals with a total privacy budget of $\rho = 0.5$. $\rho_1 = 0.45$, $c = 0.001$, $h = 1.0$.



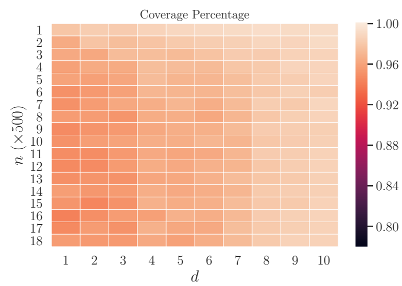
(A) ϵ-DP, objective perturbation



(B) ϵ-DP, output perturbation

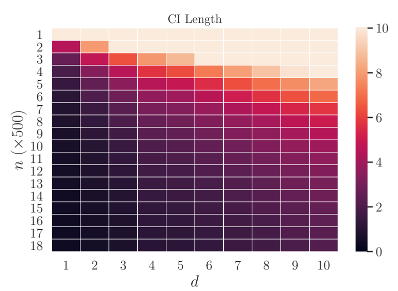


(C) zCDP, objective perturbation

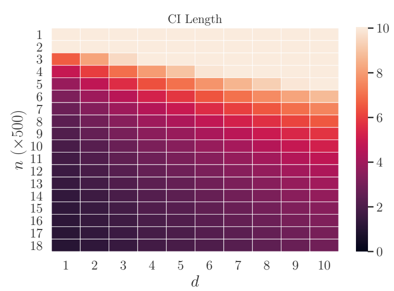


(D) zCDP, output perturbation

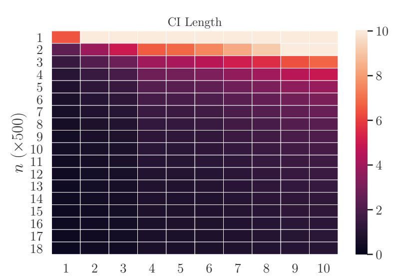
Figure 13: Coverage percentage from 1000 confidence intervals as a function of the sample size n and the dimensionality d on Adult dataset for logistic regression. $\epsilon = 1.0$, $\rho = \epsilon^2/2 = 0.5$, $c = 0.001$.



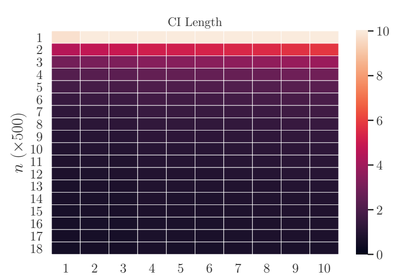
(A) ϵ-DP, objective perturbation



(B) ϵ-DP, output perturbation

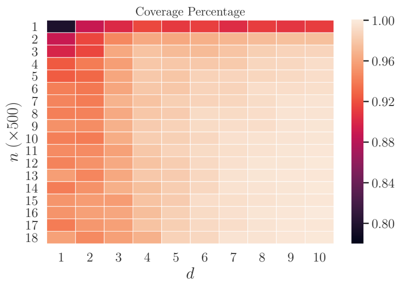


(C) zCDP, objective perturbation

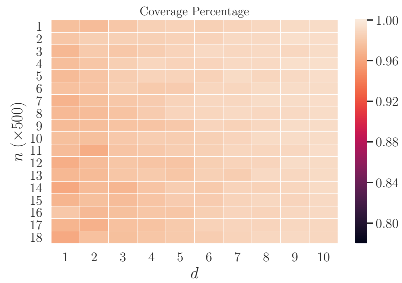


(D) zCDP, output perturbation

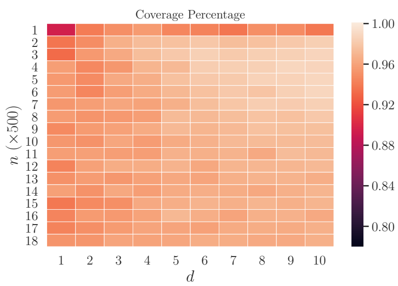
Figure 14: Average length from 1000 confidence intervals as a function of the sample size n and the dimensionality d on Adult dataset for logistic regression. $\epsilon = 1.0$, $\rho = \epsilon^2/2 = 0.5$, $c = 0.001$.



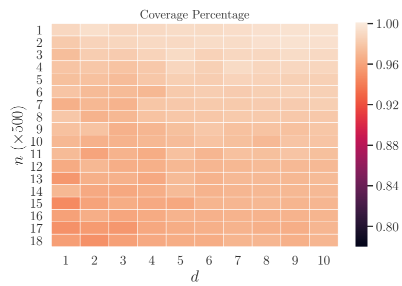
(A) ϵ-DP, objective perturbation



(B) ϵ-DP, output perturbation

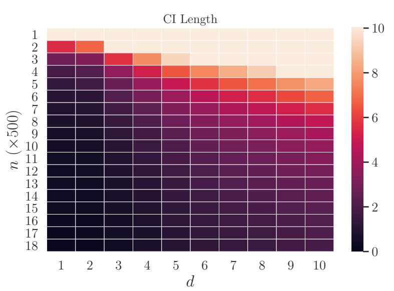


(C) zCDP, objective perturbation

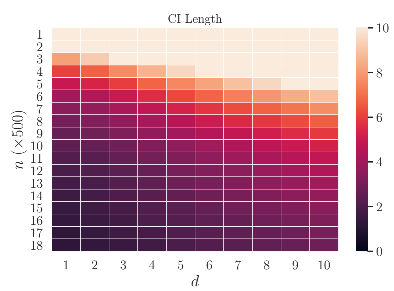


(D) zCDP, output perturbation

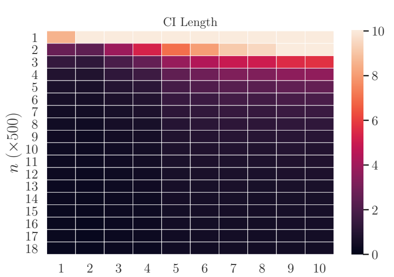
Figure 15: Coverage percentage from 1000 confidence intervals as a function of the sample size n and dimensionality d on Banking dataset for SVM. $\epsilon = 1.0$, $\rho = \epsilon^2/2 = 0.5$, $c = 0.001$, $h = 1.0$.



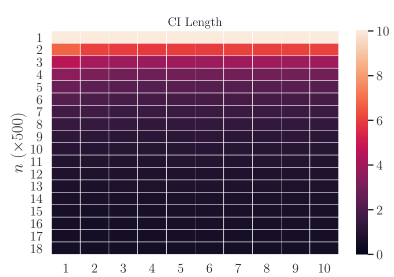
(A) ϵ-DP, objective perturbation



(B) ϵ-DP, output perturbation

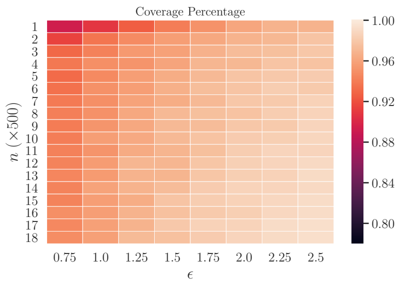


(C) zCDP, objective perturbation

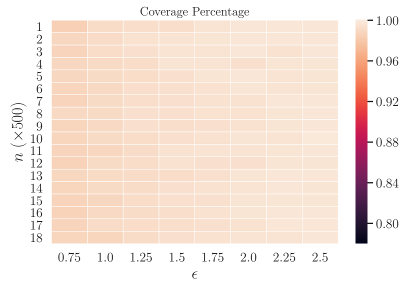


(D) zCDP, output perturbation

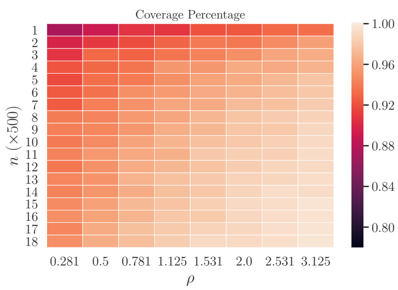
Figure 16: Average length from 1000 confidence intervals as a function of the sample size n and dimensionality d on Banking dataset for SVM. $\epsilon = 1.0$, $\rho = \epsilon^2/2 = 0.5$, $c = 0.001$, $h = 1.0$.



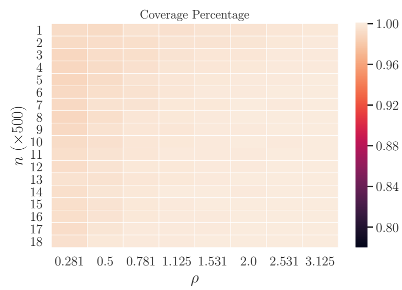
(A) ϵ-DP, objective perturbation



(B) ϵ-DP, output perturbation

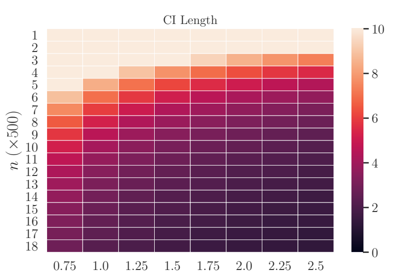


(C) zCDP, objective perturbation

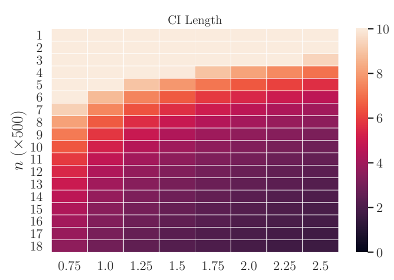


(D) zCDP, output perturbation

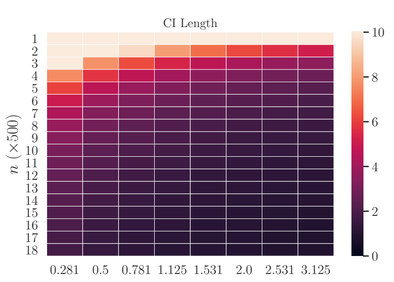
Figure 17: Coverage percentage from 1000 confidence intervals as a function of the sample size n and the total privacy budget ϵ (or ρ where $\rho = \epsilon^2/2$) on KDDCUP99 dataset for logistic regression. $d = 10$, $c = 0.001$.



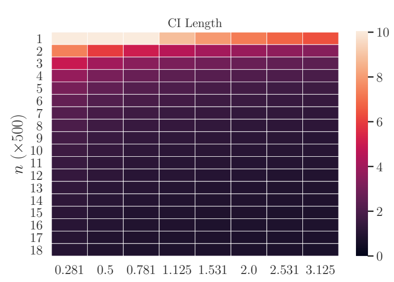
(A) ϵ-DP, objective perturbation



(B) ϵ-DP, output perturbation

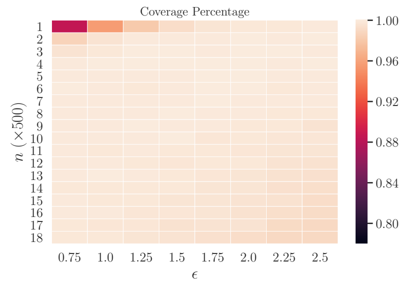


(C) zCDP, objective perturbation

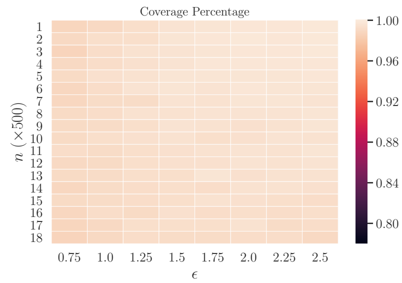


(D) zCDP, output perturbation

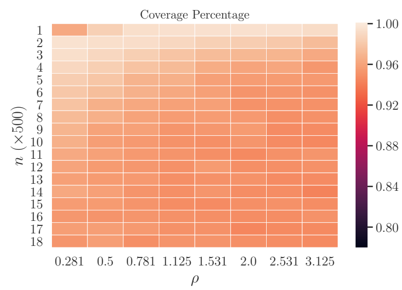
Figure 18: Average length from 1000 confidence intervals as a function of the sample size n and the total privacy budget ϵ (or ρ where $\rho = \epsilon^2/2$) on KDDCUP99 dataset for logistic regression. $d = 10$, $c = 0.001$.



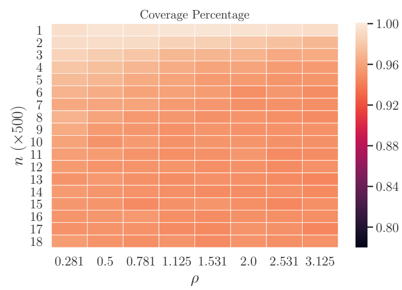
(A) ϵ-DP, objective perturbation



(B) ϵ-DP, output perturbation



(C) zCDP, objective perturbation



(D) zCDP, output perturbation

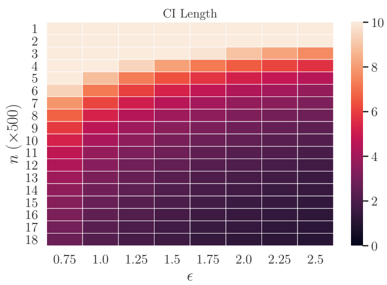
Figure 19: Coverage percentage from 1000 confidence intervals as a function of the sample size n and the total privacy budget ϵ (or ρ where $\rho = \epsilon^2/2$) on US dataset for SVM. $d = 10$, $c = 0.001$, $h = 1.0$.

B.3. **The Overhead on Sample Complexity for Differential Privacy.** See Figures 21 and 22.

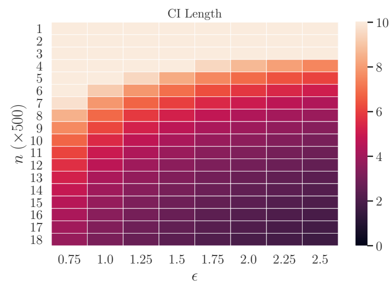
B.4. **Comparison among the Private Confidence Intervals and the Variability Intervals.** See Figures 23 to 28.

B.5. **Modeling the Relationship between Length of the Intervals and Other Parameters.** See Figures 29 to 34.

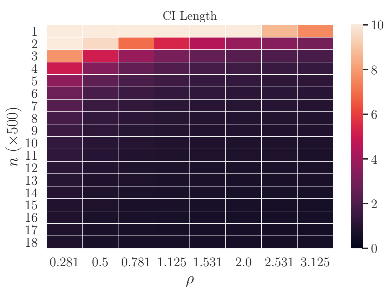
This work is licensed under the Creative Commons License Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0). To view a copy of this license, visit https://creativecommons.org/licenses/by-nc-nd/4.0/ or send a letter to Creative Commons, 171 Second St, Suite 300, San Francisco, CA 94105, USA, or Eisenacher Strasse 2, 10777 Berlin, Germany



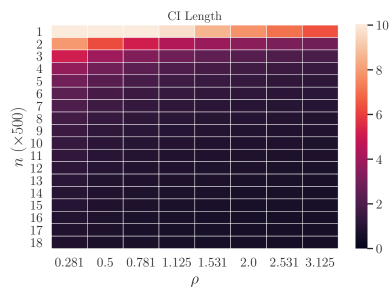
(A) ϵ-DP, objective perturbation



(B) ϵ-DP, output perturbation

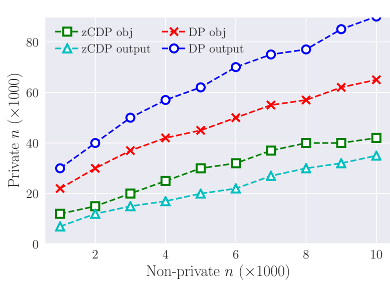


(C) zCDP, objective perturbation

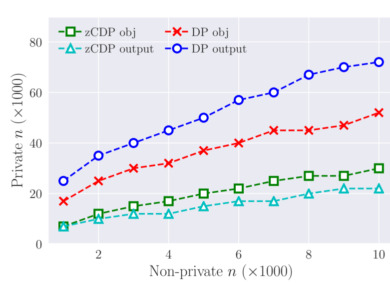


(D) zCDP, output perturbation

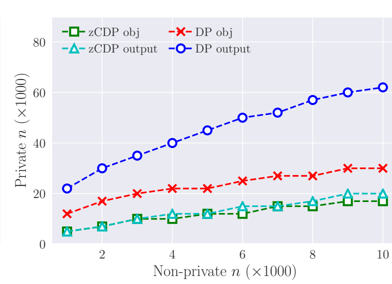
Figure 20: Average length from 1000 confidence intervals as a function of the sample size n and the total privacy budget ϵ (or ρ where $\rho = \epsilon^2/2$) on US dataset for SVM. $d = 10$, $c = 0.001$, $h = 1.0$.



(A) Adult

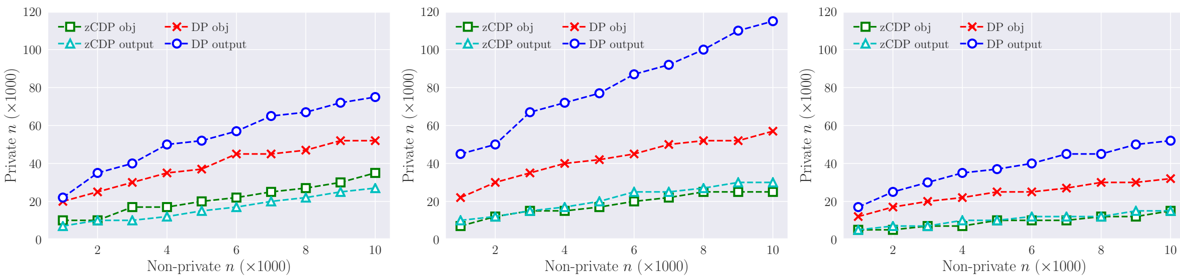


(B) Banking



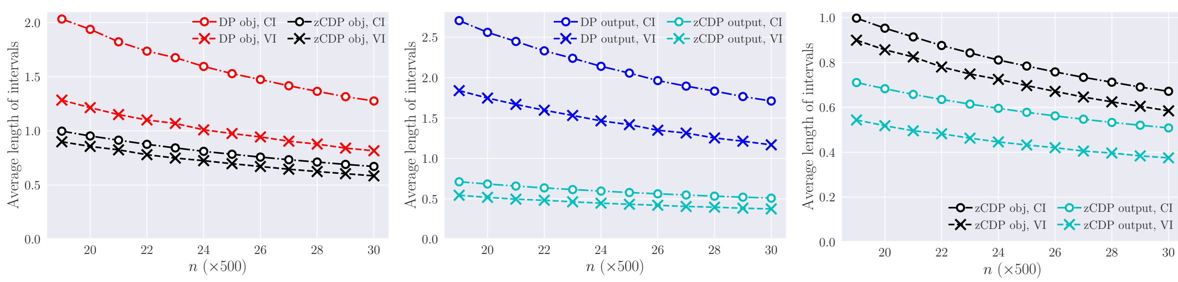
(C) US

Figure 21: The mapping between sample complexities such that the average length of the non-private confidence intervals is equivalent to that of the private confidence intervals for logistic regression. $\epsilon = 1.0$ (corresponds to $\rho = 0.5$), $d = 10$, $c = 0.001$.



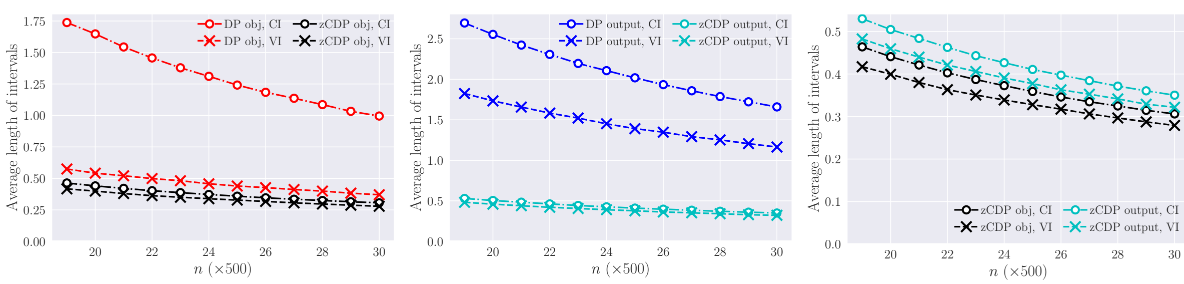
(A) Adult (B) Banking (C) US

Figure 22: The mapping between sample complexities such that the average length of the non-private confidence intervals is equivalent to that of the private confidence intervals for SVM. $\epsilon = 1.0$ (corresponds to $\rho = 0.5$), $d = 10$, $c = 0.001$, $h = 1.0$.



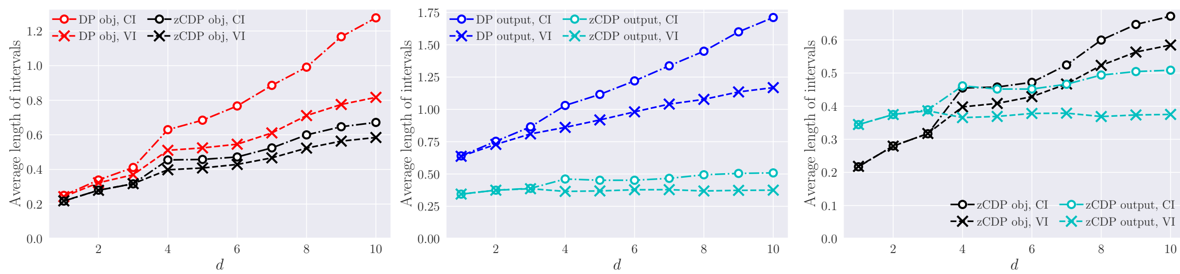
(A) ϵ-DP vs. zCDP, obj. perturbation (B) ϵ-DP vs. zCDP, output perturbation (C) obj. vs. output perturbation, zCDP

Figure 23: Comparison among length of intervals with varying n for logistic regression on Adult dataset. $d = 10$, $\epsilon = 1.0$ (corresponds to $\rho = 0.5$), $c = 0.001$.



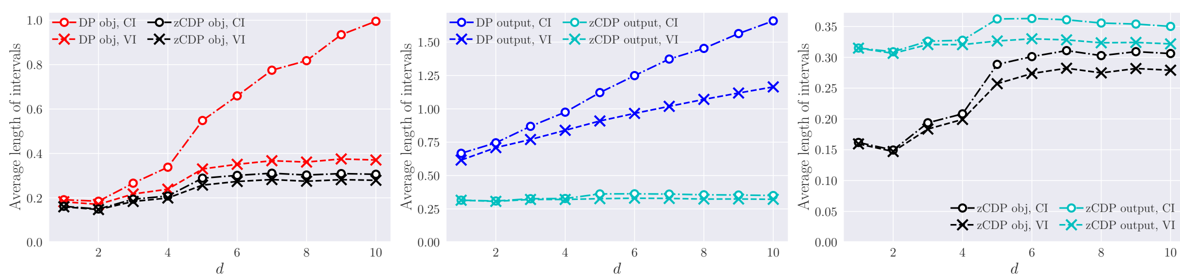
(A) ϵ-DP vs. zCDP, obj. perturbation (B) ϵ-DP vs. zCDP, output perturbation (C) obj. vs. output perturbation, zCDP

Figure 24: Comparison among length of intervals with varying n for SVM on Banking dataset. $d = 10$, $\epsilon = 1.0$ (corresponds to $\rho = 0.5$), $c = 0.001$, $h = 1.0$.



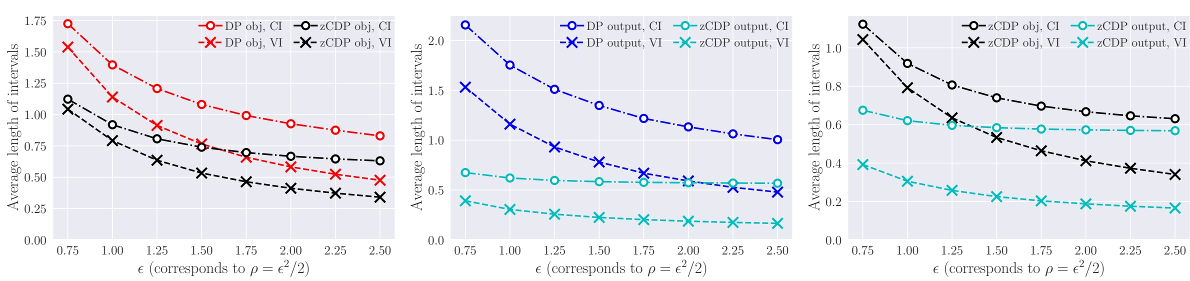
(A) ϵ-DP vs. zCDP, obj. perturbation
(B) ϵ-DP vs. zCDP, output perturbation
(C) obj. vs. output perturbation, zCDP

Figure 25: Comparison among length of intervals with varying d for logistic regression on Adult dataset. $n = 15,000$, $\epsilon = 1.0$ (corresponds to $\rho = 0.5$), $c = 0.001$.



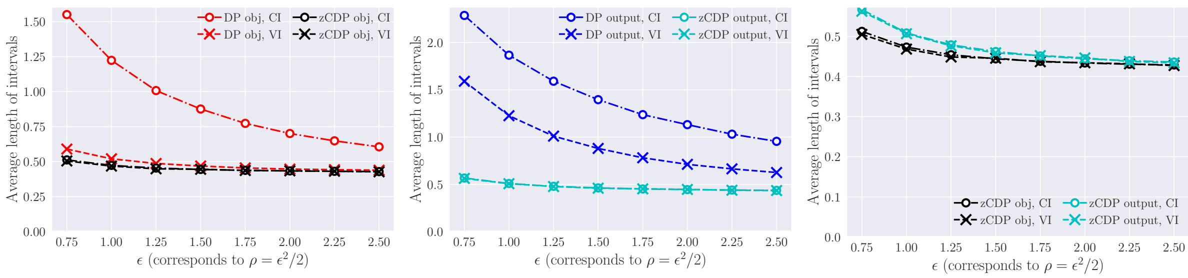
(A) ϵ-DP vs. zCDP, obj. perturbation
(B) ϵ-DP vs. zCDP, output perturbation
(C) obj. vs. output perturbation, zCDP

Figure 26: Comparison among length of intervals with varying d for SVM on Banking dataset. $n = 15,000$, $\epsilon = 1.0$ (corresponds to $\rho = 0.5$), $c = 0.001$, $h = 1.0$.



(A) ϵ-DP vs. zCDP, obj. perturbation
(B) ϵ-DP vs. zCDP, output perturbation
(C) obj. vs. output perturbation, zCDP

Figure 27: Comparison among length of intervals with varying ϵ (or ρ with $\rho = \epsilon^2/2$) for logistic regression on KDDCUP99 dataset. $n = 15,000$, $d = 10$, $c = 0.001$.

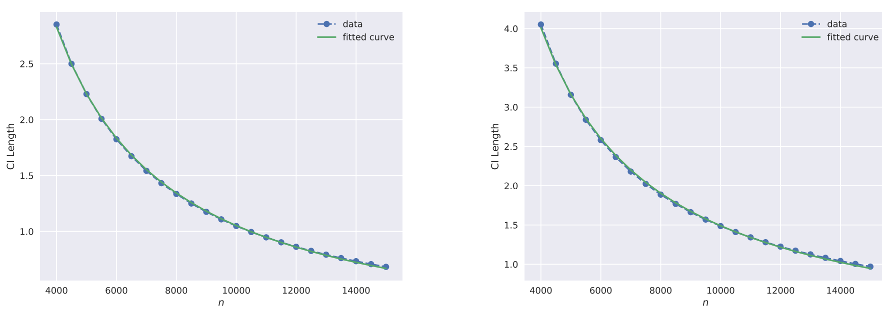


(A) ϵ-DP vs. zCDP, obj. perturbation

(B) ϵ-DP vs. zCDP, output perturbation

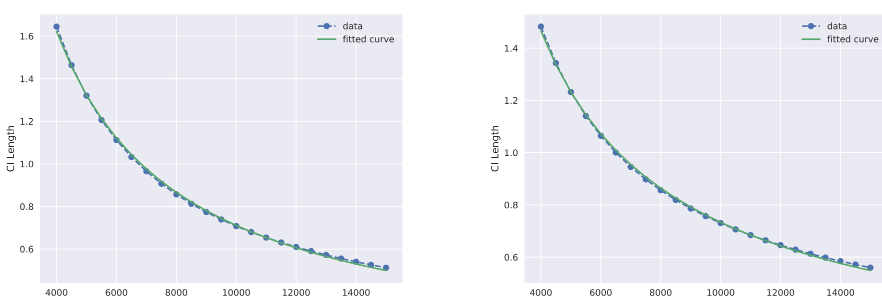
(C) obj. vs. output perturbation, zCDP

Figure 28: Comparison among length of intervals with varying ϵ (or ρ with $\rho = \epsilon^2/2$) for SVM on US dataset. $n = 15,000$, $d = 10$, $c = 0.001$, $h = 1.0$.



(A) ϵ-DP, objective perturbation, Adult

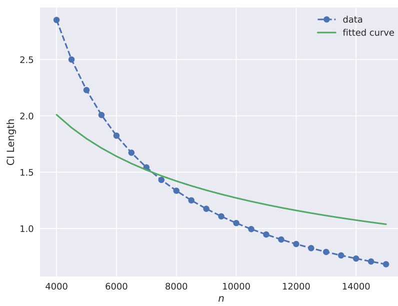
(B) ϵ-DP, output perturbation, US



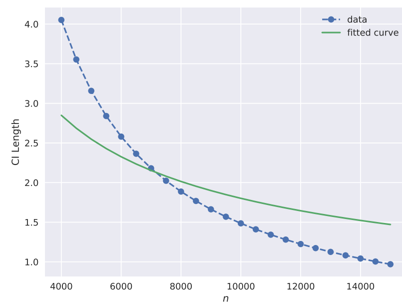
(C) zCDP, objective perturbation, Banking

(D) zCDP, output perturbation, KDDCUP99

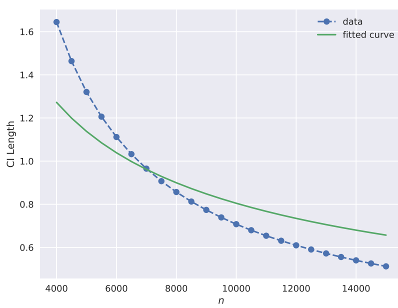
Figure 29: Relationship between average length of the confidence intervals and the sample size n for logistic regression. $d = 5$, $c = 0.001$, $\epsilon = 1.0$ (corresponds to $\rho = 0.5$). The fitted curve is $\frac{c_0}{n} + \frac{c_1}{\sqrt{n}}$.



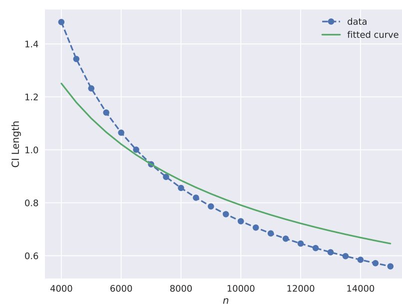
(A) ϵ-DP, objective perturbation, Adult



(B) ϵ-DP, output perturbation, US

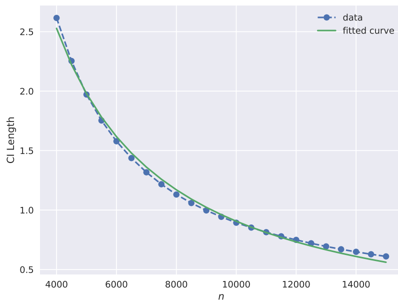


(C) zCDP, objective perturbation, Banking

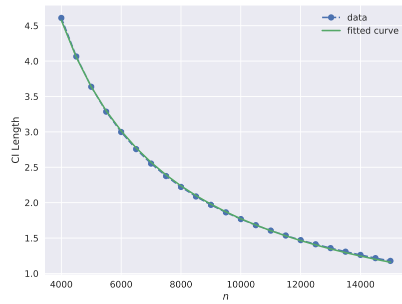


(D) zCDP, output perturbation, KDDCUP99

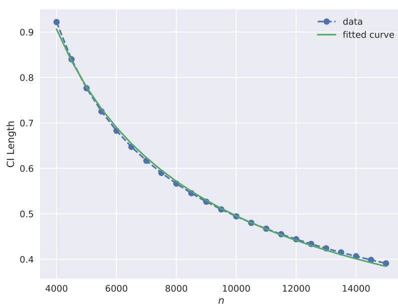
Figure 30: Relationship between average length of the confidence intervals and the sample size n for logistic regression. $d = 5$, $c = 0.001$, $\epsilon = 1.0$ (corresponds to $\rho = 0.5$). The fitted curve is $\frac{c}{\sqrt{n}}$, that shows the length for the privacy-preserving confidence intervals is not proportional to $\frac{1}{\sqrt{n}}$ as in the non-private case.



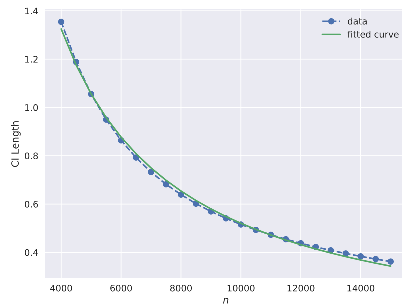
(A) ϵ-DP, objective perturbation, BR



(B) ϵ-DP, output perturbation, Adult

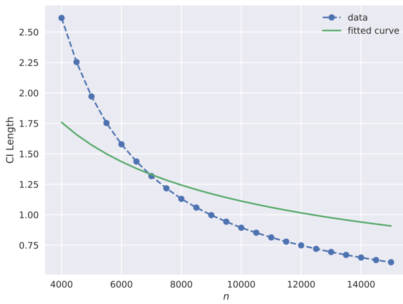


(C) zCDP, objective perturbation, US

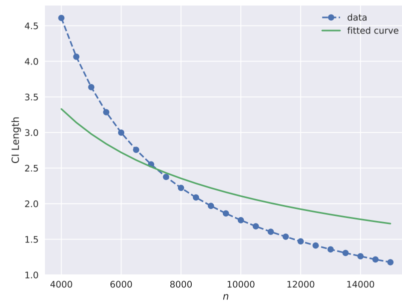


(D) zCDP, output perturbation, Banking

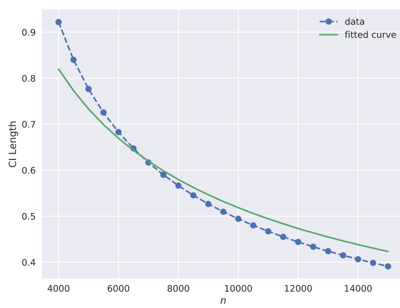
Figure 31: Relationship between average length of the confidence intervals and the sample size n for SVM. $d = 5$, $c = 0.001$, $\epsilon = 1.0$ (corresponds to $\rho = 0.5$), $h = 1.0$. The fitted curve is $\frac{c_0}{n} + \frac{c_1}{\sqrt{n}}$.



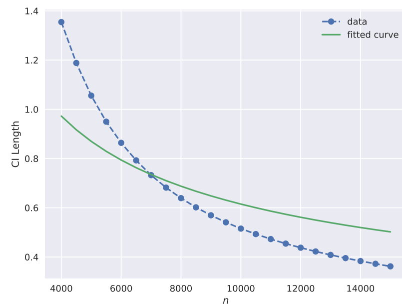
(A) ϵ-DP, objective perturbation, BR



(B) ϵ-DP, output perturbation, Adult

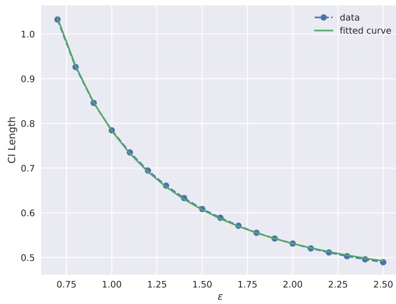


(C) zCDP, objective perturbation, US

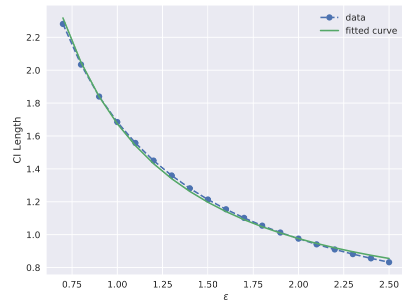


(D) zCDP, output perturbation, Banking

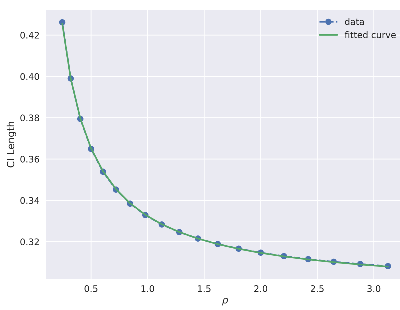
Figure 32: Relationship between average length of the confidence intervals and the sample size n for SVM. $d = 5$, $c = 0.001$, $\epsilon = 1.0$ (corresponds to $\rho = 0.5$), $h = 1.0$. The fitted curve is $\frac{c}{\sqrt{n}}$, that shows the length for the privacy-preserving confidence intervals is not proportional to $\frac{1}{\sqrt{n}}$ as in the non-private case.



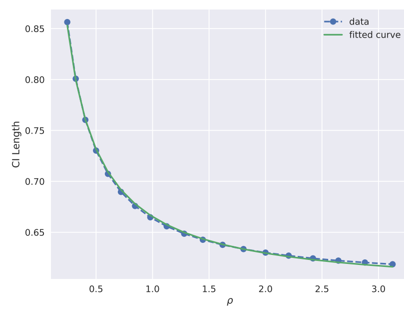
(A) ϵ-DP, objective perturbation, BR



(B) ϵ-DP, output perturbation, Adult

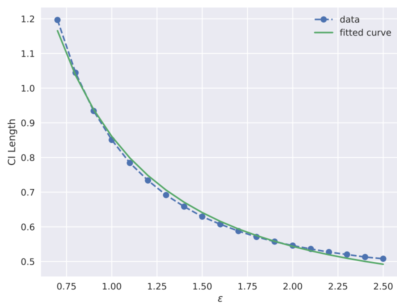


(C) zCDP, objective perturbation, US

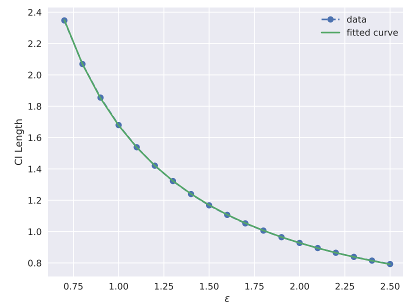


(D) zCDP, output perturbation, KDDCUP99

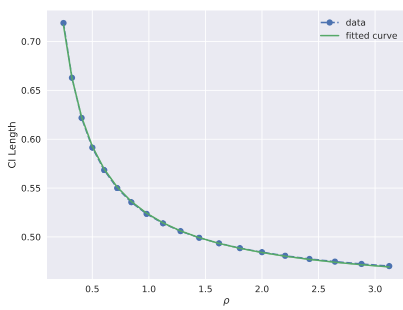
Figure 33: Relationship between average length of the confidence intervals and the total privacy budget ϵ (or $\rho = \epsilon^2/2$) for logistic regression. $n = 10000$, $d = 5$, $c = 0.001$. The fitted curve is $\sqrt{\frac{c_0}{\epsilon^2} + c_1}$ for ϵ-DP, $\sqrt{\frac{c_0}{\rho} + c_1}$ for zCDP.



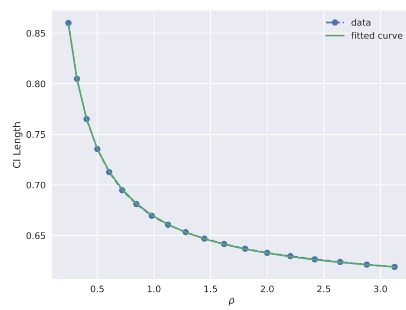
(A) ϵ-DP perturbation, objective, US



(B) ϵ-DP perturbation, output, BR



(C) zCDP, objective perturbation, Adult



(D) zCDP, output perturbation, KDDCUP99

Figure 34: Relationship between average length of the confidence intervals and the total privacy budget ϵ (or $\rho = \epsilon^2/2$) for SVM. $n = 10000$, $d = 5$, $c = 0.001$, $h = 1.0$. The fitted curve is $\sqrt{\frac{c_0}{\epsilon^2} + c_1}$ for ϵ-DP, $\sqrt{\frac{c_0}{\rho} + c_1}$ for zCDP.