Homework #0

CSE 446/546: Machine Learning
Prof. Jamie Morgenstern
Due: **Wednesday** January 11, 2023 11:59pm
38 points

Please review all homework guidance posted on the website before submitting to GradeScope. Reminders:

- Make sure to read the "What to Submit" section following each question and include all items.
- Please provide succinct answers and supporting reasoning for each question. Similarly, when discussing experimental results, concisely create tables and/or figures when appropriate to organize the experimental results. All explanations, tables, and figures for any particular part of a question must be grouped together.
- For every problem involving generating plots, please include the plots as part of your PDF submission.
- When submitting to Gradescope, please link each question from the homework in Gradescope to the location of its answer in your homework PDF. Failure to do so may result in deductions of up to 10% of the value of each question not properly linked. For instructions, see `https://www.gradescope.com/get_started#student-submission`.
- If you collaborate on this homework with others, you must indicate who you worked with on your homework by providing a complete list of collaborators on the first page of your assignment. Make sure to include the name of each collaborator, and on which problem(s) you collaborated. Failure to do so may result in accusations of plagiarism. You can review the course collaboration policy at `https://courses.cs.washington.edu/courses/cse446/23wi/assignments/`
- For every problem involving code, please include all code you have written for the problem as part of your PDF submission *in addition to* submitting your code to the separate assignment on Gradescope created for code. Not submitting all code files will lead to a deduction of up to 10% of the value of each question missing code.

Not adhering to these reminders may result in point deductions.

Changelog:

- **1/8:** Fixed typo in A10 instructions.

Probability and Statistics

A1. *[2 points]* (From Murphy Exercise 2.4.) After your yearly checkup, the doctor has bad news and good news. The bad news is that you tested positive for a serious disease, and that the test is 99% accurate (i.e., the probability of testing positive given that you have the disease is 0.99, as is the probability of testing negative given that you don't have the disease). The good news is that this is a rare disease, striking only one in 10,000 people. What are the chances that you actually have the disease?

What to Submit:

- Final Answer
- Corresponding Calculations

A2. For any two random variables X, Y the *covariance* is defined as $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$. You may assume X and Y take on a discrete values if you find that is easier to work with.

a. *[1 point]* If $\mathbb{E}[Y \mid X = x] = x$ show that $\text{Cov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$.

b. *[1 point]* If X, Y are independent show that $\text{Cov}(X, Y) = 0$.

What to Submit:

- **Parts a-b:** Proofs

A3. Let X and Y be independent random variables with PDFs given by f and g, respectively. Let h be the PDF of the random variable $Z = X + Y$.

a. *[1 point]* Show that $h(z) = \int_{-\infty}^{\infty} f(x)g(z - x)\,\mathrm{d}x$. (If you are more comfortable with discrete probabilities, you can instead derive an analogous expression for the discrete case, and then you should give a one sentence explanation as to why your expression is analogous to the continuous case.).

b. *[1 point]* If X and Y are both independent and uniformly distributed on $[0, 1]$ (i.e. $f(x) = g(x) = 1$ for $x \in [0, 1]$ and 0 otherwise) what is h, the PDF of $Z = X + Y$?

What to Submit:

- **Part a:** Proof
- **Part b:** Formula for PDF Z and corresponding calculations

A4. Let $X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2)$ be i.i.d random variables. Compute the following:

a. *[1 point]* $a \in \mathbb{R}, b \in \mathbb{R}$ such that $aX_1 + b \sim \mathcal{N}(0, 1)$.

b. *[1 point]* $\mathbb{E}[X_1 + 2X_2], \text{Var}[X_1 + 2X_2]$.

c. *[2 points]* Setting $\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, the mean and variance of $\sqrt{n}(\widehat{\mu}_n - \mu)$.

What to Submit:

- **Part a:** a, b, and the corresponding calculations
- **Part b:** $\mathbb{E}[X_1 + 2X_2]$, $\text{Var}[X_1 + 2X_2]$
- **Part c:** $\mathbb{E}[\sqrt{n}(\widehat{\mu}_n - \mu)]$, $\text{Var}[\sqrt{n}(\widehat{\mu}_n - \mu)]$
- **Parts a-c** Corresponding calculations

Linear Algebra and Vector Calculus

A5. Let $A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix}$. For each matrix A and B:

a. *[2 points]* What is its rank?

b. *[2 points]* What is a (minimal size) basis for its column span?

What to Submit:

- **Parts a-b:** Solution and corresponding calculations

A6. Let $A = \begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix}$, $b = \begin{bmatrix} -2 & -2 & -4 \end{bmatrix}^\top$, and $c = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^\top$.

a. *[1 point]* What is Ac?

b. *[2 points]* What is the solution to the linear system $Ax = b$?

What to Submit:

- **Parts a-b:** Solution and corresponding calculations

A7. For possibly non-symmetric $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$ and $c \in \mathbb{R}$, let $f(x, y) = x^\top \mathbf{A} x + y^\top \mathbf{B} x + c$. Define

$$\nabla_z f(x, y) = \begin{bmatrix} \frac{\partial f}{\partial z_1}(x, y) & \frac{\partial f}{\partial z_2}(x, y) & \dots & \frac{\partial f}{\partial z_n}(x, y) \end{bmatrix}^\top \in \mathbb{R}^n .$$

a. *[2 points]* Explicitly write out the function $f(x, y)$ in terms of the components $A_{i,j}$ and $B_{i,j}$ using appropriate summations over the indices.

b. *[2 points]* What is $\nabla_x f(x, y)$ in terms of the summations over indices *and* vector notation?

c. *[2 points]* What is $\nabla_y f(x, y)$ in terms of the summations over indices *and* vector notation?

What to Submit:

- **Part a:** Explicit formula for $f(x, y)$
- Parts b-c: Summation form and corresponding calculations
- Parts b-c: Vector form and corresponding calculations

A8. Show the following:

a. *[2 points]* Let $g : \mathbb{R} \to \mathbb{R}$ and $v, w \in \mathbb{R}^n$ such that $g(v_i) = w_i$. Find an expression for g such that $\text{diag}(v)^{-1} = \text{diag}(w)$.

b. *[2 points]* Let $\mathbf{A} \in R^{n \times n}$ be orthonormal and $x \in \mathbb{R}^n$. An orthonormal matrix is a square martix whose columns and rows are orthonormal vectors, such that $\mathbf{A}\mathbf{A}^\top = \mathbf{A}^\top \mathbf{A} = \mathbf{I}$ where $\mathbf{I}$ is the identity matrix. Show that $||\mathbf{A}x||_2^2 = ||x||_2^2$.

c. *[2 points]* Let $\mathbf{B} \in R^{n \times n}$ be invertible and symmetric. A symmetric matrix is a square matrix satisfying $\mathbf{B} = \mathbf{B}^\top$. Show that $\mathbf{B}^{-1}$ is also symmetric.

d. *[2 points]* Let $\mathbf{C} \in R^{n \times n}$ be positive semi-definite (PSD). A positive semi-definite matrix is a symmetric matrix satisfying $x^\top \mathbf{C} x \geq 0$ for any vector $x \in \mathbb{R}^n$. Show that its eigenvalues are non-negative.

What to Submit:

- **Part a:** Explicit formula for g
- **Parts a-d:** Proof

Programming

These problems are available in a .zip file, with some starter code. All coding questions in this class will have starter code. **Before attempting these problems, you will need to set up a Conda environment that you will use for every assignment in the course. Unzip the HW0-A.zip file and read the instructions in the README file to get started.**

A9. For $\nabla_x f(x, y)$ as solved for in Problem 7:

a. *[1 point]* Using native Python, implement the summation form.

b. *[1 point]* Using NumPy, implement the vector form.

c. *[1 point]* Report the difference in wall-clock time for parts a-b, and discuss reasons for the observed difference.

What to Submit:

- **Part c:** Difference in wall-clock time for parts a-b
- **Part c:** Explanation for difference (1-2 sentences)
- **Code** on Gradescope through coding submission

A10. Two random variables X and Y have equal distributions if their CDFs, F_X and F_Y, respectively, are equal, i.e. for all x, $|F_X(x) - F_Y(x)| = 0$. The central limit theorem says that the sum of k independent, zero-mean, variance $1/k$ random variables converges to a (standard) Normal distribution as k tends to infinity. We will study this phenomenon empirically (you will use the Python packages Numpy and Matplotlib). Each of the following subproblems includes a description of how the plots were generated; these have been coded for you. The code is available in the .zip file. In this problem, you will add to our implementation to explore **matplotlib** library, and how the solution depends on n and k.

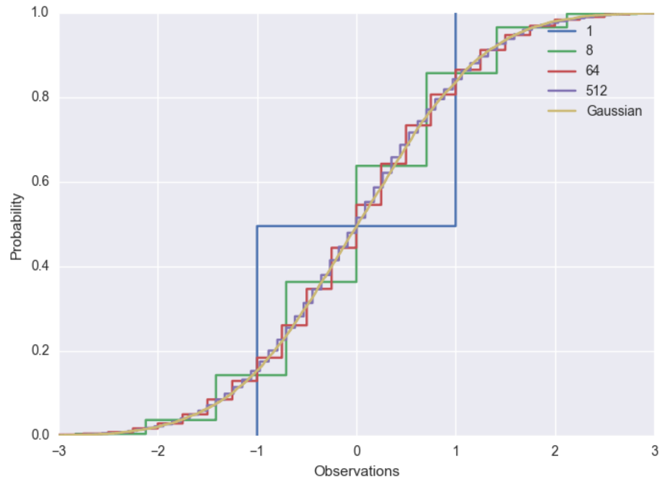
a. *[2 points]* For $i = 1, \ldots, n$ let $Z_i \sim \mathcal{N}(0, 1)$. Let $F(x)$ denote the true CDF from which each Z_i is drawn (i.e., Gaussian). Define $\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{Z_i \leq x\}$ and we will choose n large enough such that, for all $x \in \mathbb{R}$,

$$\sqrt{\mathbb{E}\left[\left(\widehat{F}_n(x) - F(x)\right)^2\right]} \leq 0.0025 \ .$$

Plot $\widehat{F}_n(x)$ from -3 to 3.

b. *[2 points]* Define $Y^{(k)} = \frac{1}{\sqrt{k}} \sum_{i=1}^{k} B_i$ where each B_i is equal to -1 and 1 with equal probability and the B_i's are independent. We know that $\frac{1}{\sqrt{k}} B_i$ is zero-mean and has variance $1/k$. For each $k \in \{1, 8, 64, 512\}$ we will generate n (same as in part a) independent copies $Y^{(k)}$ and plot their empirical CDF on the same plot as part a.

Be sure to always label your axes. Your plot should look something like the following (up to styling) (Tip: checkout `seaborn` for instantly better looking plots.)



What to Submit:

- **Part a:** Value for n (Hint: You will need to print it). **Part b:** In 1-2 sentences: How does empirical CDF change with k?
- **Parts a and b:** Plot of $\widehat{F}_n(x) \in [-3, 3]$
- **Code** on Gradescope through coding submission