Homework 0: Review Problems
CSE 446: Machine Learning

University of Washington

0 Policies [0 points]

Please read these policies. **Please answer the three questions below and include your answers marked in a "problem 0" in your solution set.** Homeworks which do not include these answers will not be graded.

Readings: Read the required material.

Submission format: Submit your report as a *single* pdf file. Also, please include all your code in the PDF file in a section at the end of your document, marked "Code"; also specify which problem(s) the code corresponds to. The report (in a single pdf file) must include all the plots and explanations for programming questions (if required). Homework solutions must be organized in order, with all plots arranged in the correct location in your submitted solutions. We highly recommend typesetting your scientific writing using LaTeX(see the website for references for free tools). Writing solutions by hand will be accepted provided they are neat; written solutions need to be scanned and included into a single pdf.

Written work: Please provide succinct answers *along with succinct reasoning for all your answers*. Points may be deducted if long answers demonstrate a lack of clarity. Similarly, when discussing the experimental results, concisely create tables and figures to organize the experimental results. In other words, all your explanations, tables, and figures for any particular part of a question must be grouped together.

Python source code: for the programming assignment. Please note that we will not accept Jupyter notebooks. Submit your code together with a neatly written README file to instruct how to run your code with different settings (if applicable). We assume that you always follow good practice of coding (commenting, structuring); these factors are not central to your grade.

Coding policies: You must write your own code. You are welcome to use any Python libraries for data munging, visualization, and numerical linear algebra. Examples includes Numpy, Pandas, and Matplotlib. You may **not**, however, use any machine learning libraries such as Scikit-Learn, TensorFlow, or PyTorch, unless explicitly specified for that question. If in doubt, post to the message boards.

Collaboration: For homework 0, it is encouraged you make an attempt to solve each non-programming question on your own before you discuss questions with other students (HW0 is meant as refresher to help you out downstream). It is acceptable for you to discuss problems with other students; it is not acceptable for students to look at another students written answers. It is acceptable for you to discuss coding questions with others; it is not acceptable for students to look at another students code. Each student must understand, write, and hand in their own answers. In addition, each student must write and submit their own code in the programming part of the assignment.

Acknowledgments: We expect the students not to refer to or seek out solutions in published material from previous years, on the web, or from other textbooks. Students are certainly encouraged to read extra material for a deeper understanding.

Extra Credit Policy: Before you do the extra credit problems, do all the regular questions. Extra credit points will only be awarded if there are (honest attempts at) answers to *all* the regular questions. This is because they are not designed to be alternative questions to the regular questions.

0.1 List of Collaborators

List the names of all people you have collaborated with and for which question(s). **For homework** 0**, you must make an attempt to solve each non-programming question entirely on your own before you discuss questions with other students.** This problem set is meant as a refresher and doing it individually is for your benefit.

0.2 List of Acknowledgements

If you do inadvertently find an assignment's answer, acknowledge for which question and provide an appropriate citation (there is no penalty, provided you include the acknowledgement). If not, then write "none".

0.3 Certify that you have read the instructions

Please make sure to read and follow these instructions. Write "I have read and understood these policies" to certify this.

1 Probability and Statistics

1. *[3 points]* (independence and dependence) Let X and Y be real independent random variables with PDFs given by f and g, respectively. Let h be the PDF of the random variable $Z = X + Y$. Show that $h(z) = \int_{-\infty}^{\infty} f(x)g(z-x)dx$. (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.).
2. (conditional probabilities) Suppose 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). Let h be the PDF of the random variable $Z = X + Y$. For these given explicit distributions,

   (a) *[3 points]* What is h?
   (b) *[3 points]* What is $\mathbb{P}\left(X \leq \frac{1}{2} \mid X + Y \geq \frac{5}{4}\right)$?

3. *[3 points]* (change of variable) A random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ is Gaussian distributed with mean μ and variance σ^2. Given that for any $a, b \in \mathbb{R}$, we have that $Y = aX + b$ is also Gaussian, find a, b such that $Y \sim \mathcal{N}(0, 1)$.
4. 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) *[3 points]* If $\mathbb{E}[Y|X = x] = x$ show that $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])^2]$.
   (b) *[3 points]* If X, Y are independent show that $\text{Cov}(X, Y) = 0$.

5. If $f(x)$ is a PDF, the cumulative distribution function (CDF) is defined as $F(x) = \int_{-\infty}^{x} f(y)dy$. For any function $g : \mathbb{R} \to \mathbb{R}$ and random variable X with PDF $f(x)$, recall that the expected value of $g(X)$ is defined as $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(y)f(y)dy$. For a boolean event A, define $\mathbf{1}\{A\}$ as 1 if A is true, and 0 otherwise. Thus, $\mathbf{1}\{x \leq a\}$ is 1 whenever $x \leq a$ and 0 whenever $x > a$. Note that $F(x) = \mathbb{E}[\mathbf{1}\{X \leq x\}]$. Let $X_1, \ldots, X_n$ be *independent and identically distributed* random variables with CDF $F(x)$. Define $\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i \leq x\}$. Note, for every x, that $\widehat{F}_n(x)$ is an *empirical estimate* of $F(x)$.

   (a) *[3 points]* For any x, what is $\mathbb{E}[\widehat{F}_n(x)]$?
   (b) *[3 points]* For any x, the variance of $\widehat{F}_1(x)$ is $\mathbb{E}[(\widehat{F}_1(x) - F(x))^2]$. Show that $\text{Variance}(\widehat{F}_1(x)) = F(x)(1 - F(x))$.
   (c) *[6 points]* For any x, the variance of $\widehat{F}_n(x)$ is $\mathbb{E}[(\widehat{F}_n(x) - F(x))^2]$. Show that $\text{Variance}(\widehat{F}_n(x)) = \frac{F(x)(1-F(x))}{n}$.
   (d) *[6 points]* Using your answer to c, show that for all $x \in \mathbb{R}$, we have $\mathbb{E}[(\widehat{F}_n(x) - F(x))^2] \leq \frac{1}{4n}$.

2 Geometry and Linear Algebra

1. (Hyperlanes) Assume w is n-dimensional vector and b is a scalar. A hyper-plane in $\mathbb{R}^n$ is the set $\{x : x \in \mathbb{R}^n, \text{ s.t. } w^\top x + b = 0\}$.

(a) *[3 points]* ($n = 2$ example) Draw the hyperplane for $w = [-1,\ 2]^\top$, $b = 2$? Label your axes.

(b) *[3 points]* ($n = 3$ example) Draw the hyperplane for $w = [1,\ 1,\ 1]^\top$, $b = 0$? Label your axes.

(c) *[6 points]* (distance) Given some $x_0 \in \mathbb{R}^n$, find the *squared* distance to the hyperlane defined by $w^\top x + b = 0$. In other words, solve the following optimization problem:

$$\begin{aligned} &\min_x ||x_0 - x||^2 \\ &\text{s.t. } \ w^\top x + b = 0 \end{aligned}$$

Remember, we want the squared distance not the closest x.

2. *[6 points]* (Rank) Let $A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 2 \end{bmatrix}$?

(a) What is the rank of A? Why?

(b) What is a (minimal size) basis for the column span?

3. *[6 points]* (Linear Equations and Matrix Multiplication) Let $A = \begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix}$, $b = [-2,\ -2,\ -4]^\top$, and $c = [1,\ 1,\ 1]^\top$.

(a) What is Ac?

(b) What is the solution to the linear system $Ax = b$?

4. *[3 points]* (Linear Algebra) For possibly non-symmetric $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n \times n}$ and $c \in \mathbb{R}$, define $f(x, y) = x^\top \boldsymbol{A} x + y^\top \boldsymbol{B} x + c$ for $x, y \in \mathbb{R}^n$. Explicitly write out the function $f(x, y)$ in terms of the components $\boldsymbol{A}_{i,j}$ and $\boldsymbol{B}_{i,j}$ using appropriate summations over the indices.

3 Programming

1. *[6 points]* For the A and b as defined in Problem 2.3, use numpy to compute:

(a) What is A^{-1}?

(b) What is $A^{-1}b$? What is Ac? Take a screen shot of your answer.

2. *[15 points]* 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 goes off to infinity. We will study this phenomenon empirically (you will use the Python packages Numpy and Matplotlib). 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. It is easy to verify (you should) that $\frac{1}{\sqrt{k}} B_i$ is zero-mean and has variance $1/k$.

(a) For $i = 1, \ldots, n$ let $Z_i \sim \mathcal{N}(0, 1)$. If $F(x)$ is the true CDF from which each Z_i is drawn (i.e., Gaussian) and $\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{Z_i \leq x)$, use the answer to problem 1.5 above to choose n large enough such that, for all $x \in \mathbb{R}$, $\sqrt{\mathbb{E}[(\widehat{F}_n(x) - F(x))^2]} \leq 0.0025$, and plot $\widehat{F}_n(x)$ from

-3 to 3.
(Hint: use `Z=numpy.random.randn(n)` to generate the random variables, and `import matplotlib.pyplot as plt;`
`plt.step(sorted(Z), np.arange(1,n+1)/float(n))` to plot).

(b) For each $k \in \{1, 8, 64, 512\}$ generate n independent copies $Y^{(k)}$ and plot their empirical CDF on the same plot as part a.
(Hint: `np.sum(np.sign(np.random.randn(n, k))*np.sqrt(1./k), axis=1)` generates n of the $Y^{(k)}$ random variables.)

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