

# Homework #0

CSE 446/546: Machine Learning  
Prof. Jamie Morgenstern and Simon Du  
Due: **Wednesday** October 6, 2021 11:59pm  
**A:** 38 points, **B:** 7 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 *[5 points]*. For instructions, see [https://www.gradescope.com/get\\_started#student-submission](https://www.gradescope.com/get_started#student-submission).
- Please recall that B problems, indicated in boxed text, are only graded for 546 students, and that they will be weighted at most 0.2 of your final GPA (see website for details). In Gradescope there is a place to submit solutions to A and B problems separately. You are welcome to create just a single PDF that contains answers to both, submit the same PDF twice, but associate the answers with the individual questions in Gradescope.
- If you collaborate on this homework with others, you must indicate who you worked with on your homework. Failure to do so may result in accusations of plagiarism.
- For every problem involving code, please include the code as part of your PDF for the 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 *[1 point]*.

Not adhering to these reminders may result in point deductions.

## 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?

Bayes Rule:

$$P(D|+) = \frac{P(+|D) \cdot P(D)}{P(+)}$$

where  $D$  is you have the disease and  $+$  is you test positive. So  $P(D|+)$  is probability you actually have the disease given you test positive. We solve for this. We know  $P(+|D) = 0.99$  (the probability that that you test positive given you have the disease).  $P(D) = \frac{1}{10,000}$  is the probability that you have the disease. And,  $P(+)$  (probability that you test positive) is

$$P(+) = P(+|D) \cdot P(D) + P(+|\neg D) \cdot P(\neg D)$$

$$P(+) = 0.99 \cdot 0.0001 + P(+|\neg D) \cdot 0.9999$$

So,

$$P(D|+) = \frac{P(+|D) \cdot P(D)}{P(+|D) \cdot P(D) + P(+|\neg D) \cdot P(\neg D)}$$

$$P(D|+) = \frac{0.99 \cdot 0.0001}{0.99 \cdot 0.0001 + 0.01 \cdot 0.9999}$$

$$P(D|+) = \frac{0.99 \cdot 0.0001}{0.99 \cdot 0.0001 + 0.01 \cdot 0.9999}$$

$$P(D|+) = 0.00980392156862745$$

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  $\mathbb{E}[Y | X = x] = x$  show that  $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])^2]$ .

*Proof.* Let  $\mu = \mathbb{E}[X]$  Then we want to show that  $\text{Cov}(X, Y) = \mathbb{E}[(X - \mu)^2]$ . Recall,

$$\begin{aligned}\text{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[(X - \mathbb{E}[X])^2] \\ \text{Cov}(X, Y) &= \mathbb{E}[XY] - \mathbb{E}[X] \mathbb{E}[Y]\end{aligned}$$

Let's find  $\mathbb{E}[XY]$  and  $\mathbb{E}[X] \mathbb{E}[Y]$  separately using the law of iterative Expectations.

$$\begin{aligned}\mathbb{E}[Y] &= \mathbb{E}[\mathbb{E}[Y | X = x]] \\ &= \int \mathbb{E}[Y | X = x] f(x) \, dx \\ &= \int x f(x) \, dx \\ &= \mathbb{E}[X]\end{aligned}$$

$$\begin{aligned}\mathbb{E}[XY] &= \mathbb{E}[\mathbb{E}[XY | X = x]] \\ &= \mathbb{E}[X \mathbb{E}[Y | X = x]] \\ &= \int X^2 f(x) \, dx \\ &= \mathbb{E}[X^2]\end{aligned}$$

Plugging this back into the Cov equation above we get

$$\begin{aligned}\text{Cov}(X, Y) &= \mathbb{E}[XY] - \mathbb{E}[X] \mathbb{E}[Y] \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ &= \text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]\end{aligned}$$

□

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

*Proof.* By definition,

$$\begin{aligned}\text{Cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY - \mathbb{E}[Y]X - \mathbb{E}[Y]X + \mathbb{E}[X]\mathbb{E}[Y]] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[\mathbb{E}[Y]X] + \mathbb{E}[\mathbb{E}[X]\mathbb{E}[Y]] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[Y]\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]\end{aligned}$$

If  $X$  and  $Y$  are independent then  $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$  so  $\text{Cov}(X, Y) = 0$ . □

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  $h(z) = \int_{-\infty}^{\infty} f(x)g(z-x) dx$ . (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.).

*Proof.* Let's start with some definitions:  $h(z)$  is the PDF which is equivalent to  $h(z) = \frac{d}{dz}H(z)$  where  $H(z)$  is the CDF. Now, let's find the CDF  $H(z)$ .

$$\begin{aligned} H(z) &= P(Z \leq z_0) \\ &= P(X + Y \leq z_0) \end{aligned}$$

Let  $X$  be specified as  $x$  then  $Y \leq z - x$ .

$$\begin{aligned} H(z) &= P(X + Y \leq z) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f(x, y) dy dx \\ &= \int_{-\infty}^{\infty} f(x) \int_{-\infty}^{z-x} g(y) dy dx \text{ since } x \text{ and } y \text{ are independent} \\ &= \int_{-\infty}^{\infty} f(x) \int_{-\infty}^{z-x} g(y) dy dx \text{ to evaluate the } y \text{ variable integral} \\ &= \int_{-\infty}^{\infty} f(x)G(z-x) dx \end{aligned}$$

Then, we plug that into the CDF to get the PDF and using the Leibniz rule, we get

$$h(z) = \int_{-\infty}^{\infty} f(x)g(z-x) dx$$

□

- 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  $Z = X + Y$ ?

Since  $X \in [0, 1]$  and  $Y \in [0, 1]$  then  $X + Y = Z \in [0, 2]$ . The pdf is

$$\begin{aligned} h(z) &= \int_{-\infty}^{\infty} f(x)g(z-x) \, dx \\ &= \int_0^1 f(x)g(z-x) \, dx \\ &= \int_0^1 g(z-x) \, dx \end{aligned}$$

$z-x$  is non-zero only for  $0 \leq z-x \leq 1$ , or  $z-1 \leq x \leq z$ .

$$\begin{aligned} h(z) &= \begin{cases} \int_0^z \, dx & \text{for } z \in [0, 1] \\ \int_{z-1}^1 \, dx & \text{for } z \in [1, 2] \end{cases} \\ h(z) &= \begin{cases} 1 & \text{for } z \in [0, 1] \\ 2-z & \text{for } z \in [1, 2] \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

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

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

Let  $y = aX_1 + b \sim \mathcal{N}(0, 1)$  so  $0 = a\mu + b$  and  $1 = a^2\sigma^2$ . Rearranging, we get one such  $a, b$ :

$$a = \frac{1}{\sigma}$$

$$b = -a\mu = -\frac{\mu}{\sigma}$$

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

$\mathbb{E}[X_1 + 2X_2] = \mathbb{E}[X_1] + 2\mathbb{E}[X_2]$  by linearity. Since  $X_1$  and  $X_2$  are i.i.d random variables, the expectation is just  $\mu$ . So,  $= \mu + 2\mu = 3\mu$ .

$$\begin{aligned}\text{Var}[X_1 + 2X_2] &= \mathbb{E}[(X_1 + 2X_2)^2] - \mathbb{E}[X_1 + 2X_2]^2 \\ &= \mathbb{E}[X_1^2 + 4X_1X_2 + 4X_2^2] - (3\mu)^2 \\ &= \mathbb{E}[X_1^2] + \mathbb{E}[4X_1X_2] + \mathbb{E}[4X_2^2] - 9\mu^2\end{aligned}$$

Since  $\mathbb{E}[X_1^2] = \text{Var}[X_1] + \mathbb{E}[X_1]^2 = \sigma^2 + \mu^2$  then,

$$\begin{aligned}\text{Var}[X_1 + 2X_2] &= \sigma^2 + \mu^2 + 4\mathbb{E}[X_1]\mathbb{E}[X_2] + \mathbb{E}[4X_2^2] - 9\mu^2 \\ &= \sigma^2 + \mu^2 + 4\mu^2 + 4(\mu^2 + \sigma^2) - 9\mu^2 \\ &= 5\sigma^2\end{aligned}$$



- c. *[2 points]* Setting  $\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n X_i$ , the mean and variance of  $\sqrt{n}(\hat{\mu}_n - \mu)$ .  
Let's find  $\mathbb{E}[\hat{\mu}_n]$  first.

$$\begin{aligned}\mathbb{E}[\hat{\mu}_n] &= \mathbb{E}\left[\frac{1}{n} \sum_i^n X_i\right] \\ &= \frac{1}{n} \sum_i^n \mathbb{E}[X]_i \\ &= \frac{1}{n} \sum_i^n \mu_i \\ &= \frac{1}{n} n\mu = \mu\end{aligned}$$

So then,  $\sqrt{n}(\hat{\mu}_n - \mu) = \sqrt{n}(\mu - \mu) = 0$ .

## 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?

A's rank is two because the sum of the second and third columns are a multiple of the first. B also has a rank of two because the first and second column added up to the third.

- b. *[2 points]* What is a (minimal size) basis for its column span?  
Same as the rank - the basis is the minimal size to get unique vectors or column span.

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

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

$$\begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0+2+4 \\ 2+4+2 \\ 3+3+1 \end{bmatrix} = \begin{bmatrix} 6 \\ 8 \\ 7 \end{bmatrix}$$

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

$$\begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \\ -4 \end{bmatrix}$$

We can find the solution by using row reduced method (R2 refers to the the second row, similarly for R1 and R3):

$$\left[ \begin{array}{ccc|c} 0 & 2 & 4 & -2 \\ 2 & 4 & 2 & -2 \\ 3 & 3 & 1 & -4 \end{array} \right]$$

Swap R1 and R2:

$$\left[ \begin{array}{ccc|c} 2 & 4 & 2 & -2 \\ 0 & 2 & 4 & -2 \\ 3 & 3 & 1 & -4 \end{array} \right]$$

$R3 = R3 + (-\frac{3}{2})R1$ :

$$\left[ \begin{array}{ccc|c} 2 & 4 & 2 & -2 \\ 0 & 2 & 4 & -2 \\ 0 & -3 & -2 & -1 \end{array} \right]$$

$R3 = R3 + (\frac{3}{2})R2$ :

$$\left[ \begin{array}{ccc|c} 2 & 4 & 2 & -2 \\ 0 & 2 & 4 & -2 \\ 0 & 0 & 4 & -4 \end{array} \right]$$

$R2 = \frac{R2}{2}$  and  $R3 = \frac{R3}{4}$

$$\left[ \begin{array}{ccc|c} 2 & 4 & 2 & -2 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 1 & -1 \end{array} \right]$$

$R2 = R2 - 2R3$

$$\left[ \begin{array}{ccc|c} 2 & 4 & 2 & -2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{array} \right]$$

$R1 = R1 - 2R2 - R3$

$$\left[ \begin{array}{ccc|c} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{array} \right]$$

So,

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix}$$

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

$$\nabla_z f(x, y) = \left[ \frac{\partial f}{\partial z_1}(x, y) \quad \frac{\partial f}{\partial z_2}(x, y) \quad \dots \quad \frac{\partial f}{\partial z_n}(x, y) \right]^\top.$$

- 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.

$f(x, y) = x^\top \mathbf{A}x + y^\top \mathbf{B}x + c$  in summation form is 
$$f(x, y) = \sum_i^n \sum_j^n x_i a_{ij} x_j + \sum_i^n \sum_j^n y_i b_{ij} x_j + c$$

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

$$\nabla_x f(x, y) = \begin{bmatrix} \sum_i^n x_i a_{1i} + \sum_i^n x_i a_{i1} + \sum_i^n y_i b_{i1} \\ \cdot \\ \cdot \\ \sum_i^n x_i a_{ni} + \sum_i^n x_i a_{in} + \sum_i^n y_i b_{in} \end{bmatrix}$$

Vector form follows from this to become,

$$\nabla_x f(x, y) = \begin{bmatrix} \sum_i^n x_i a_{1i} \\ \cdot \\ \cdot \\ \sum_i^n x_i a_{ni} \end{bmatrix} + \begin{bmatrix} \sum_i^n x_i a_{i1} \\ \cdot \\ \cdot \\ \sum_i^n x_i a_{in} \end{bmatrix} + \begin{bmatrix} \sum_i^n y_i b_{i1} \\ \cdot \\ \cdot \\ \sum_i^n y_i b_{in} \end{bmatrix}$$

$$\boxed{\nabla_x f(x, y) = Ax + A^\top x + B^\top y}$$

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

$$\nabla_y f(x, y) = \begin{bmatrix} \sum_i^n b_{1i} x_i \\ \cdot \\ \sum_i^n b_{1i} x_i \end{bmatrix}$$

which is the following in vector notation:

$$\boxed{\nabla_y f(x, y) = Bx}$$



A8. Show the following:

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

Given  $g(v_i) = w_i$ , we can find an expression for  $g$  that satisfies  $\text{diag}(v)^{-1} = \text{diag}(w)$ . Let's start with the definition of diagonal matrices and inverse matrix.

$$\begin{aligned}\text{diag}(v) &= \begin{bmatrix} v_1 & & 0 \\ & \ddots & \\ 0 & & v_n \end{bmatrix} \\ \text{diag}(v)^{-1} &= \begin{bmatrix} \frac{1}{v_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{v_n} \end{bmatrix} \\ \text{diag}(w) &= \begin{bmatrix} w_1 & & 0 \\ & \ddots & \\ 0 & & w_n \end{bmatrix}\end{aligned}$$

Since  $\text{diag}(v)^{-1} = \text{diag}(w)$ ,

$$\begin{aligned}\text{diag}(v)^{-1} &= \text{diag}(w) \\ \begin{bmatrix} \frac{1}{v_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{v_n} \end{bmatrix} &= \begin{bmatrix} w_1 & & 0 \\ & \ddots & \\ 0 & & w_n \end{bmatrix}\end{aligned}$$

Substitute  $g(v_i) = w_i$ ,

$$\begin{bmatrix} \frac{1}{v_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{v_n} \end{bmatrix} = \begin{bmatrix} g(v_1) & & 0 \\ & \ddots & \\ 0 & & g(v_n) \end{bmatrix}$$

Thus,

$$\boxed{g(v_i) = \frac{1}{v_i}}$$

.

b. [2 points] Let  $\mathbf{A} \in \mathbb{R}^{n \times n}$  be orthonormal and  $x \in \mathbb{R}^n$ . Show that  $\|\mathbf{A}x\|_2^2 = \|x\|_2^2$ .

*Proof.*  $\mathbf{A}$  being orthonormal means that each basis is orthogonal to each other and of unit length. We will show  $\|\mathbf{A}x\|_2^2 = \|x\|_2^2$ . Let's start with the right hand side (RHS),

$$\begin{aligned}\|x\|_2^2 &= \langle x, x \rangle && \text{by definition} \\ &= x^\top x\end{aligned}$$

Now the left hand side (LHS) where we will use the fact that  $A^\top A = \mathbf{I}$  since  $\mathbf{A}$  is orthonormal.

$$\begin{aligned}\|\mathbf{A}x\|_2^2 &= (Ax)^\top (Ax) \\ &= x^\top A^\top Ax \\ &= x^\top (A^\top A)x \\ &= x^\top x\end{aligned}$$

Thus, RHS = LHS. □

c. [2 points] Let  $\mathbf{B} \in \mathbb{R}^{n \times n}$  be invertible and symmetric. Show that  $\mathbf{B}^{-1}$  is also symmetric.

*Proof.* Since  $\mathbf{B}$  is invertible,  $\mathbf{B}^{-1}$  exists. We will show that  $\mathbf{B}^{-1}$  is symmetric given that  $\mathbf{B}$  is symmetric. Since  $\mathbf{B}$  is symmetric, there is some matrix  $X$  and a diagonal matrix  $D$  such that  $\mathbf{B} = XDX^\top$ . Then,

$$\begin{aligned}\mathbf{B}^{-1} &= (XDX^\top)^{-1} \\ &= X^{-\top} D^{-1} X^{-1}\end{aligned}$$

$X$  is an invertible matrix and the inverse of a diagonal matrix is another diagonal matrix. Let  $Y = X^{-\top}$  and  $D_0$  is a diagonal matrix.

$$\mathbf{B}^{-1} = YD_0Y^\top$$

Thus,  $\mathbf{B}^{-1}$  is also symmetric. □

- d. [2 points] Let  $\mathbf{C} \in \mathbb{R}^{n \times n}$  be positive semi-definite (PSD). Show that its eigenvalues are non-negative.

*Proof.* By definition of PSD,  $\mathbf{C}v = \lambda v$  for all vectors  $v \neq 0$ . Then,  $v^\top \mathbf{C}v = v^\top \lambda v$  where the left hand side (LHS) is non-negative since  $C$  is positive and  $v^\top v$  must be non-negative. Then the right hand side (RHS) means that  $\lambda$  must also be non-negative since  $v^\top \lambda v = \lambda v^\top v = \lambda \geq 0$ .  $\square$

## Programming

**These problem is available in zip file**, with some starter code. All coding questions in this class will have starter code. **Before attempting these problem make sure your coding environment is working.** See instructions in README file in zip on course website.

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

- [1 point] Using native Python implement the summation form.
- [1 point] Using NumPy implement the vector form.
- [1 point] Report the difference in wall-clock time for parts a-b and discuss reasons for this discrepancy (if relevant).

| n    | vanilla   | numpy    |
|------|-----------|----------|
| 20   | 0.181 ms  | 0.214 ms |
| 200  | 20.122 ms | 0.553 ms |
| 500  | 99 ms     | 0.808 ms |
| 1000 | 375 ms    | 1.304 ms |

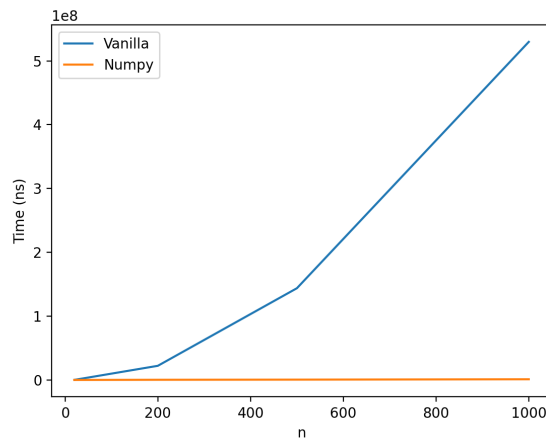


Figure 1: Wall time vs.  $n$  (size of vector/matrix) for vanilla and numpy methods.

Clearly, Figure 1 shows there is a large discrepancy between the vanilla and the numpy methods to get  $\nabla_x f(x, y)$ . As  $n$  increases the computation time for vanilla also increases exponentially, but numpy remains relatively constant. This is because numpy performs the operations on the array as a whole while vanilla is running through each individual element in the list one at a time. As a rule of thumb, built-in python functions are better and faster than one you make yourself (especially with for loops).

A10. [4 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  $k$  tends 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. We know that  $\frac{1}{\sqrt{k}} B_i$  is zero-mean and has variance  $1/k$ .

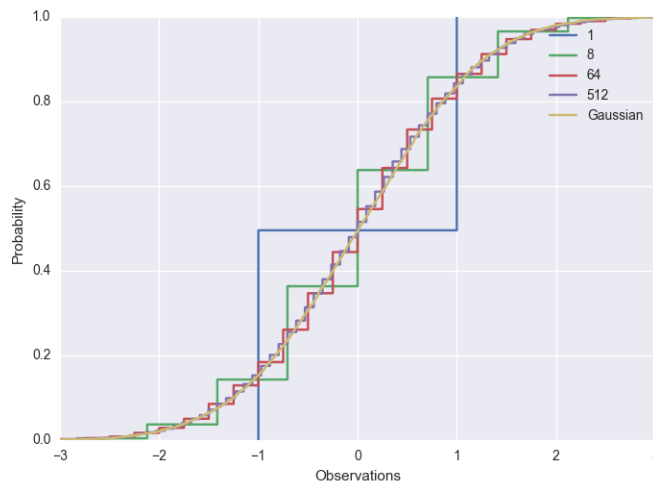
- Note that below are descriptions of how the plots are generated. We coded this part up and it is available in zip file. In this problem is meant you will explore matplotlib library, and then explain how solution changes with  $k$  in part c.
- For  $i = 1, \dots, n$  let  $Z_i \sim \mathcal{N}(0, 1)$ . If  $F(x)$  is the true CDF from which each  $Z_i$  is drawn (i.e., Gaussian) and  $\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{Z_i \leq x\}$ , we will choose  $n$  large enough such that, for all  $x \in \mathbb{R}$ ,

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

and plot  $\hat{F}_n(x)$  from  $-3$  to  $3$ .

- For each  $k \in \{1, 8, 64, 512\}$  we will generate  $n$  (same as in part D) independent copies  $Y^{(k)}$  and plot their empirical CDF on the same plot as part b.

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.)



## Solution

- Part b:** Value for  $n$  is 40000.
- Parts a and c:** In 1-2 sentences: How does empirical CDF change with  $k$ ?  
The CDF more closely approximates the continuous solution as  $k$  increases.
- Parts b and c:** Plot of  $\hat{F}_n(x) \in [-3, 3]$

