Homework 0

Stats 315A, Winter 2026

January 5, 2026

Due date: Monday January 12 at 11:59pm on Canvas.

The primary purpose of this assignment is to provide a refresher of some linear algebra and statistics at a level that you should be comfortable with in order to succeed in this course. In addition, question 7 relates to decision theory introduced in lecture 1, and question 8 is a Pytorch primer.

Question 1 (4 points): Let $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times k}$, $C \in \mathbb{R}^{k\times k}$, and $D \in \mathbb{R}^{k\times n}$. Assume that A and C are invertible. Define the $(n+k)\times(n+k)$ matrix

$$M := \begin{bmatrix} A & B \\ -D & C^{-1} \end{bmatrix}.$$

(a) Show that if $X \in \mathbb{R}^{n\times n}$ and $Y \in \mathbb{R}^{k\times n}$ solve

$$M \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} I \\ 0 \end{bmatrix},$$

where I is the $n \times n$ identity, then $X = (A + BCD)^{-1}$.

(b) Use this to demonstrate the Woodbury matrix identity that

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}.$$

so long as $C^{-1} + DA^{-1}B$ is invertible.

(c) In the case that $k = 1$, so that $B = b$ and $D = d^T$ for n-vectors b and d and $C = c$ is a scalar, give a (slightly) simplified form for the above. Give a sufficient condition for $(A + bcd^T)$ to be invertible.

Question 2 (Symmetric Matrices, **4 points**): Let A be a symmetric matrix. We say A is positive definite, denoted $A \succ 0$, if $x^T Ax > 0$ for all $x \neq 0$. We say $A \succeq 0$, meaning A is positive semidefinite, if $x^T Ax \geq 0$ for all x.

(a) Show that if $A \succ 0$, then A is full rank.

(b) Show that B has linearly independent columns if and only if $B^T B$ is positive definite.

(c) We write $A \succeq B$ if $A - B \succeq 0$. Show that if U is any matrix of appropriate size, then $A + UU^T \succeq A$.

Question 3 (5 points): The *operator norm* of a matrix C is defined by

$$\|C\|_{\text{op}} := \max_{\|u\|_2=1, \|v\|_2=1} u^T Cv = \max_{\|u\|_2\leq 1, \|v\|_2\leq 1} u^T Cv. \tag{0.1}$$

(a) If $D = \mathrm{diag}(d_1, \ldots, d_n)$, argue that $\|D\|_{\mathrm{op}} = \|d\|_\infty = \max_i |d_i|$. *Hint.* Cauchy-Schwarz.

(b) Let $A = U\Sigma V^T$ be the singular value decomposition of A. Show that $\|A\|_{\mathrm{op}} = \max_i |\sigma_i(A)|$, where $\sigma_i(A)$ are the singular values of A. *Hint.* First show that if Q has orthonormal columns, then $\|Q^T u\|_2 \le 1$ for any unit vector u.

(c) If A is symmetric with eigenvalues $\lambda_1(A) \ge \cdots \ge \lambda_n(A)$, why is $\|A\|_{\mathrm{op}} = \max\{|\lambda_1(A)|, |\lambda_n(A)|\}$?

Question 4 (Multivariate normal, **2 points**): Let $X \sim \mathsf{N}(\mu_1, \Sigma_1)$ and $Y \sim \mathsf{N}(\mu_2, \Sigma_2)$ be independent multivariate normals. Give the distribution of

$$Z = AX + BY.$$

(You do not need to prove your answer is correct.)

Question 5 (**5 points**): Let $X_i \overset{\text{iid}}{\sim} \mathsf{N}(\mu, \sigma^2)$, $i = 1, \ldots, n$. Define the sample mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ and the leave-one-out sample means $\overline{X}_{-i} = \frac{1}{n-1}\sum_{j \ne i} X_j$. Give the distribution of the vector $Z \in \mathbb{R}^n$ with entries

$$Z_i := \overline{X}_{-i} - X_i.$$

Hint. Write $Z = AX$ for a matrix A, where $X = [X_i]_{i=1}^n$.

Question 6 (EM as a minorization algorithm, Exercise 8.7 from Hastie et al. (2009), **5 points**): A function $g(x, y)$ is said to *minorize* a function $f(x)$ if

$$g(x, y) \le f(x), \qquad g(x, x) = f(x),$$

for all x, y in the domain.

This notion is useful for maximizing $f(x)$, since it is easy to show that $f(x)$ is non-decreasing under the update

$$x^{(s+1)} = \arg\max_x g(x, x^{(s)}).$$

There are analogous definitions for *majorization*, which apply to the minimization of a function $f(x)$. The resulting class of algorithms are known as *MM algorithms*, standing for either "Minorize–Maximize" or "Majorize–Minimize."

Show that the EM algorithm (Hastie et al., 2009, Section 8.5.2)[1] is an example of an MM algorithm, using

$$Q(\theta', \theta) + \log \Pr(Z \mid \theta) - Q(\theta, \theta)$$

to minorize the observed-data log-likelihood $\ell(\theta'; Z)$. (Note that only the first term involves the relevant parameter θ'.)

Question 7 (Focal loss, **3 points**): This question relates to the *focal loss* of Lin et al. (2017). Consider multi-class classification with response $y \in \{1, 2, \ldots, K\}$ and prediction $t \in \mathbb{R}^K$. The focal loss is

$$\ell_F(t, y) = -(1 - p_y)^\gamma \log p_y, \text{ for } p_y = \exp\{t_y\} / \sum_k \exp\{t_k\}.$$

(a) Show that the logistic loss for binary classification from class, $\ell_L(f, \tilde{y}) = \log(1 + \exp\{-f\tilde{y}\})$, coincides with the focal loss with $\gamma = 0$ when we reparameterize the logistic regression predictor as $t = [t_1, t_2] = [0, f]$ and consider response $y = \tilde{y} + 1 \in \{1, 2\}$ rather than $\tilde{y} \in \{0, 1\}$.

[1] You may wish to refer to this section for a refresher on the EM algorithm.

(b) Both the focal loss and the logistic loss are larger when the probability predicted for the true class (p_y) is smaller. But when $\gamma \geq 0$, compared to ℓ_L, ℓ_F will place a greater relative loss (or *focus*) on an example with small p_y vs. an example with a $p'_y > p_y$. Show that this is true by comparing the ratios of the losses.

Question 8 (Pytorch Primer, **20 points**): For assignments in this course it will be helpful use Google Colab notebooks and pytorch. Complete the pytorch primer at:
https://colab.research.google.com/drive/1CQ0JidZwKQaPaL8stDoT1tCq0udgxncM.

Follow the instructions provided at the end of the notebook to convert your solutions to a PDF, and upload this PDF to Canvas.

Question 9 (Time and collaboration, **2 points**):

(a) How many hours in total did you spend on this assignment? (This will help to calibrate future assignments.)

(b) With whom (if anyone) did you collaborate on this assignment?

References

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer, New York, 2 edition, 2009.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, 2017.