Engr108: Introduction to Matrix Methods
October 2023

B. Ayazifar and J. Duchi

Midterm Examination

This is an 80 minute in-class midterm examination.

You may use one pieces of paper (single-sided), but otherwise this examination is closed book. While you may ask us questions if you find a question confusing, we've tried pretty hard to make the exam unambiguous and clear, so we're unlikely to say much.

Please respect the honor code.

All problems have equal weight. Some are (quite) straightforward. Others, not so much.

Some problems involve applications. But you do not need to know *anything* about the problem area to solve the problem; the problem statement contains everything you need.

The problems do *not* appear in order of increasing difficulty.

Name: ______________________________

Stanford ID #: ______________________________

M.1 Consider the vector

$$x = \begin{bmatrix} 4 \\ 3 \end{bmatrix}.$$

You may tackle each of the various parts of this problem independently of the others, and in any order you wish.

(a) Determine a *nonzero* vector $u \in \mathbf{R}^2$ such that $\langle x, u \rangle = \|x\| \, \|u\|$. There is no unique answer, but you must show your work.

(b) Determine a *nonzero* vector $v \in \mathbf{R}^2$ such that $\langle x, v \rangle = 0$. There is no unique answer, but you must show your work.

(c) Determine a *nonzero* vector $y \in \mathbf{R}^2$ whose distance from x is 5—that is,

$$\mathbf{dist}(x, y) = \|x - y\| = 5.$$

There is no unique answer, but you must show your work.

M.2 We construct a block vector

$$z = \begin{bmatrix} x \\ y \end{bmatrix}$$

from the vectors

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \in \mathbf{R}^m \qquad \text{and} \qquad y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in \mathbf{R}^n.$$

The mean μ_z of the block vector z can be expressed in terms of the respective means μ_x and μ_y of the vectors x and y according to the equation

$$\mu_z = \alpha\, \mu_x + \beta\, \mu_y,$$

where α and β are real coefficients. Determine a simple expression for each of α and β in terms of the dimension parameters m and n.

M.3 A common strategy in k means is to normalize the N data vectors $x_1, \ldots, x_N \in \mathbf{R}^n$ to make their scales comparable, setting $\widetilde{x}_1 = x_1 / \|x_1\|, \ldots, \widetilde{x}_N = x_N / \|x_N\|$.

(a) What is the norm $\|\widetilde{x}_j\|$?

(b) Suppose that the k-means algorithm has converged to give cluster centers $z_1, \ldots, z_k$. Assuming the $\widetilde{x}_i$ are all distinct, when would $\|z_1\| = 1$? (Recall that if G_1 is the set of indices assigned to cluster 1, then $z_1 = \frac{1}{|G_1|} \sum_{i \in G_1} \widetilde{x}_i$.) Justify your answer. *Hint.* It may be useful to drawn a picture in dimension $n = 2$.

M.4 Here we investigate a few consequences of the definitions of correlations. Recall that for n vectors x and y, we define the de-meaned vectors

$$\widetilde{x} = x - \mathbf{avg}(x)\mathbf{1} \quad \text{and} \quad \widetilde{y} = y - \mathbf{avg}(y)\mathbf{1}$$

and the associated correlation coefficient

$$\rho_{xy} := \frac{\langle \widetilde{x}, \widetilde{y} \rangle}{\|\widetilde{x}\| \, \|\widetilde{y}\|}.$$

Throughout, assume that x and y are *not* scalar multiples of the all-ones vector $\mathbf{1}$ and $n \geq 2$.

(a) Suppose the entries of y are an *affine* function of those of x, that is, y has the form $y = \alpha x + \beta \mathbf{1}$, where α and β are (real) scalars. Explain why

(i) it must be the case that $\alpha \neq 0$ (in one short sentence)

(ii) if $\mu_x = \mathbf{avg}(x)$ and $\mu_y = \mathbf{avg}(y)$, it must be the case that

$$\beta = \mu_y - \alpha \mu_x.$$

(b) If $y = \alpha x + \beta \mathbf{1}$ as in the preceding part show that

$$\rho_{xy} = \begin{cases} -1 & \text{if } \alpha < 0 \\ 1 & \text{if } \alpha > 0. \end{cases}$$

Suppose now and for the remainder of the problem that $n = 2$, that is, $x = (x_1, x_2)$ and $y = (y_1, y_2)$ are 2-vectors, and that x and y each have distinct entries $x_1 \neq x_2$ and $y_1 \neq y_2$.

(c) True or false? Any such pair x, y can be expressed as $y = \alpha x + \beta \mathbf{1}$. Provide a succinct, yet clear and convincing explanation.

(d) Show that ρ_{xy} is either $+1$ or -1.

M.5 Extra credit. An engineer is choosing between systems to predict a signal $x \in \mathbf{R}^n$. For a prediction $p \in \mathbf{R}^n$ of the signal's values at times $i = 1, 2, \ldots, n$, the engineer considers three measures: the first takes a threshold $t > 0$ and evaluates

$$L_0(x, p) = t \cdot \frac{1}{n} \sum_{i=1}^{n} 1\{|x_i - p_i| \geq t\},$$

where $1\{A\} = 1$ if its argument is true and 0 otherwise. The second is

$$L_1(x, p) = \frac{1}{n} \sum_{i=1}^{n} |x_i - p_i|,$$

and the third

$$L_2(x, p) = \mathbf{rms}(x - p) = \sqrt{\frac{1}{n} \|x - p\|^2}.$$

The engineer claims that

$$L_0(x, p) \overset{(a)}{\leq} L_1(x, p) \overset{(b)}{\leq} L_2(x, p).$$

(a) Justify inequality (a).

(b) Justify inequality (b) by a judicious use of Cauchy-Schwarz. *Hint.* Consider the vector $v \in \mathbf{R}^n$ with entries $v_i = \frac{1}{n}|x_i - p_i|$.

(c) There are m different possible prediction systems, which produce prediction vectors $p^1, p^2, \ldots, p^m \in \mathbf{R}^n$. The engineer will choose one of the losses L_k, and for the choice $k \in \{0, 1, 2\}$, will use the system $j^\star$ satisfying

$$j^\star = \operatorname*{argmin}_{j \in \{1,\ldots,m\}} L_k(x, p^j).$$

The engineer *only* cares that the predictions satisfy $|x_i - p_i| < .5$ the majority of the time. Which loss most reflects the engineer's preferences? Why?