COS 598 – Machine Learning - Homework #1

Due: 5:00PM 02/12/2021

Homework Submission: Homeworks must be submitted via Brightspace as pdf files. This includes your code when appropriate. Please use a high quality scanner if possible, as found at the library or your departmental copy room. If you must use your phone, please don't just take photos, at least use an app like CamScanner that provides some correction for shading and projective transformations.

1) **Linear Algebra (20 pts).**

(a) (10 pts) Show that if U is an orthogonal matrix, then for all $\boldsymbol{x} \in \mathbb{R}^d$, $\|\boldsymbol{x}\| = \|U\boldsymbol{x}\|$, where the norm is the Euclidean norm.

(b) (10 pts) Show that all 2×2 orthogonal matrices have the form

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$$

for some θ. Give a geometric interpretation of the effect of these two transformations.

2) **Probability (20 pts).** Let random variables X and Y be jointly distributed random variables. Further assume they are jointly continuous with joint pdf $p(x, y)$. Show the following results.

(i) $\mathbb{V}[X] = \mathbb{E}[X^2] - [\mathbb{E}[X]]^2$, where $\mathbb{V}[X]$ is the variance of random variable X.

(ii) If X and Y are independent then $\mathbb{E}[(X+1)Y] = \mathbb{E}[Y]\,(\mathbb{E}[X]+1)$

(iii) If X and Y take values in $\{0, 1\}$ and covariance X and Y is zero, $COV(X, Y) = 0$, then X and Y are independent.

3) **Positive (semi-)definite matrices (20 pts).** Let A be a real, symmetric $d \times d$ matrix. We say A is *positive semi-definite* (PSD) if, for all $\boldsymbol{x} \in \mathbb{R}^d$, $\boldsymbol{x}^T A\boldsymbol{x} \geq 0$. We say A is *positive definite* (PD) if, for all $\boldsymbol{x} \neq \mathbf{0}$, $\boldsymbol{x}^T A\boldsymbol{x} > 0$. We write $A \succeq 0$ when A is PSD, and $A \succ 0$ when A is PD.

The *spectral theorem* (which we will assume without proof) says that every real symmetric matrix A can be expressed via the *spectral decomposition*

$$A = U\Lambda U^T,$$

where U is a $d\times d$ matrix such that $UU^T = U^TU = I$ (called an orthogonal matrix), and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_d)$. Multiplying on the right by U we see that $AU = U\Lambda$. If we let $\mathbf{u}_i$ denote the i-th column of U, we have $A\mathbf{u}_i = \lambda_i\mathbf{u}_i$ for each i. This expression reveals that the λ_i are eigenvalues of A, and the corresponding columns are eigenvectors associated to λ_i. The eigenvalues constitute the "spectrum" of A, and the spectral decomposition is also called the eigenvalue decomposition of A.

Using the spectral decomposition, show that

(a) (10 pts) A is PSD iff $\lambda_i \geq 0$ for each i.

(b) (10 pts) A is PD iff $\lambda_i > 0$ for each i.

Hint: Use the identity

$$U \Lambda U^T = \sum_{i=1}^{d} \lambda_i \mathbf{u}_i \mathbf{u}_i^T,$$

which can be verified just by showing that the matrices representing the left and right hand sides have the same entries.