Linear Algebra and Calculus Exercises: Part II

CSCI 567 Machine Learning

Fall 2022

Instructor: Vatsal Sharan

MULTIPLE-CHOICE QUESTIONS: one or more correct choices for each question.

Q1 Which of the following statements are true? PSD stands for positive semi-definite.

(a) $XX^\top$ is a PSD matrix if and only if X is PSD.

(b) If X and Y are PSD matrices, then so is $\lambda X + \mu Y$ for any $\lambda, \mu \in \mathbb{R}$.

(c) If $X - Y$ and $X + Y$ are PSD matrices, then so are X and Y.

(d) All eigenvalues of a symmetric PSD matrix are non-negative.

Q2 Suppose A and B are two positive definite matrices. Which matrix may NOT be positive definite?

(a) A^{-1}

(b) $A + B$

(c) $AA^\top$

(d) $A - B$

SHORT-ANSWER QUESTION. *The following questions use linear algebra and calculus in ML formulations. They particularly test your knowledge of gradients of multivariate functions.*

Q3 Consider the following optimization problem:

$$\boldsymbol{w}_* = \arg\min_{\boldsymbol{w} \in \mathbb{R}^d} \|\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}\|_2^2 + \boldsymbol{w}^T \boldsymbol{M} \boldsymbol{w}$$

Here, $\boldsymbol{X} \in \mathbb{R}^{n \times d}$, $\boldsymbol{y} \in \mathbb{R}^n$, $\boldsymbol{M} \in \mathbb{R}^{d \times d}$ is a positive definite matrix and $\|\cdot\|_2$ stands for the ℓ_2 norm. Find the closed form solution for $\boldsymbol{w}_*$. (This optimization problem is a generalization of ℓ_2 *regularization*, which we will see in class.)

Q3 Assume we have a training set $(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n) \in \mathbb{R}^d \times \mathbb{R}$, where each outcome y_i is generated by a probabilistic model $\boldsymbol{w}_*^{\mathrm{T}} \boldsymbol{x}_i + \epsilon_i$ with ϵ_i being an independent Gaussian noise with zero-mean and variance σ^2 for some $\sigma > 0$. In other words, the probability of seeing any outcome $y \in \mathbb{R}$ given $\boldsymbol{x}_i$ is

$$\Pr(y \mid \boldsymbol{x}_i; \boldsymbol{w}_*, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(y - \boldsymbol{w}_*^{\mathrm{T}} \boldsymbol{x}_i)^2}{2\sigma^2}\right).$$

Assume σ is fixed and given, find the maximum likelihood estimation for $\boldsymbol{w}_*$. In other words, first write down the probability of seeing the outcomes $y_1, \ldots, y_n$ given $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ as a function of the value of $\boldsymbol{w}_*$; then find the value of $\boldsymbol{w}_*$ that maximizes this probability. You can assume $\boldsymbol{X}^{\mathrm{T}} \boldsymbol{X}$ is invertible where $\boldsymbol{X}$ is the data matrix with each row corresponding to the feature of an example.