HW1

CISC684010 - Fall 2021

Due Date: September 10th at 11PM

1 Linear Algebra (10 pts)

Singular value decomposition (SVD) factorizes a $m \times n$ matrix X as $X = U\Sigma V^T$, where $U \in \mathbb{R}^{m\times m}$ and $U^TU = UU^T = I$, $\Sigma \in \mathbb{R}^{m\times n}$ contains non-increasing non-negative values along its diagonal and zeros elsewhere, and $V \in \mathbb{R}^{n\times n}$ and $V^TV = VV^T = I$. Given the SVD of a matrix $X = U\Sigma V^T$, what is the eigendecomposition of XX^T? (You need to define an appropriate square matrix Q and diagonal matrix Λ such that $XX^T = Q\Lambda Q^{-1}$.)

2 Positive (Semi-)Definite Matrices (10 pts each part)

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

The *spectral theorem* says that every real symmetric matrix A can be expressed $A = U\Lambda U^T$, where U is a $d \times d$ matrix such that $UU^T = U^TU = I$ (called an orthogonal matrix), and $\Lambda = diag(\lambda_1, ..., \lambda_d)$. 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 $\mathbf{u}_i$ are eigenvectors associated to λ_i.

Using the spectral decomposition, show that

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

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

Hint: $U\Lambda U^T = \sum_{i=1}^{d} \lambda_i \mathbf{u}_i \mathbf{u}_i^T$ (you do not need to prove this. But please try to understand why this holds.)

3 Maximum Likelihood Estimation (15 pts)

Consider a random variable $\mathbf{X}$ (possibly a vector) whose distribution (density function or mass function) belongs to a parametric family. The density or mass function may be written $f(\mathbf{x}; \theta)$, where θ is called the parameter, and can be either a scalar or vector. For example, in the Gaussian family, θ can be a two-dimensional vector consisting of the mean and variance. Suppose the parametric family is known, but the value of the parameter is unknown. It is often of interest to estimate this parameter from observations of $\mathbf{X}$.

Maximum likelihood estimation is one of the most important parameter estimation techniques. Let $\mathbf{X_1}, ..., \mathbf{X_n}$ be i.i.d (independent and identically distributed) random variables distributed according to $f(\mathbf{x}; \theta)$. By independence, the joint distribution of the observations is the product

$$\prod_{i=1}^{n} f(\mathbf{X_i}; \theta) \tag{1}$$

Viewed as a function of θ, this quantity is called the likelihood of θ. It is often more convenient to work with the *log-likelihood*,

$$\sum_{i=1}^{n} \log f(\mathbf{X_i}; \theta) \tag{2}$$

A maximum likelihood estimate (MLE) of θ is any parameter

$$\hat{\theta} \in \arg\max_{\theta} \sum_{i=1}^{n} \log f(\mathbf{X_i}; \theta) \tag{3}$$

where "arg max" denotes the set of all values achieving the maximum. If there is a unique maximizer, it is called the maximum likelihood estimate. Let $\mathbf{X_1}, ..., \mathbf{X_n}$ be iid Poisson random variables with intensity parameter λ. Determine the maximum likelihood estimator of λ.

4 Unconstrained Optimization (10 pts each part)

In this problem you will prove some of properties of unconstrained optimiziation problems. For the next two parts, the following fact will be helpul. A twice continuously differentiable function admits the quadratic expansion [1]

$$f(\mathbf{x}) = f(\mathbf{y}) + \langle \bigtriangledown f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle + \frac{1}{2} \langle \mathbf{x} - \mathbf{y}, \bigtriangledown^2 f(\mathbf{y})(\mathbf{x} - \mathbf{y}) \rangle + o(\|\mathbf{x} - \mathbf{y}\|^2) \tag{4}$$

where $o(t)$ denotes a function satisfying $\lim_{t \to 0} \frac{o(t)}{t} = 0$, as well as the expansion

$$f(\mathbf{x}) = f(\mathbf{y}) + \langle \bigtriangledown f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle + \frac{1}{2} \langle \mathbf{x} - \mathbf{y}, \bigtriangledown^2 f(\mathbf{y} + t(\mathbf{x} - \mathbf{y}))(\mathbf{x} - \mathbf{y}) \rangle \tag{5}$$

for some $t \in (0, 1)$.

a) Show that if f is twice continuously differentiable and $\mathbf{x}^*$ is a local minimizer, then $\bigtriangledown^2 f(\mathbf{x}^*) \succeq 0$, *i.e.*, the Hessian of f is positive semi-definite at the local minimizer $\mathbf{x}^*$.

[1] $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y}$ is the inner product of $\mathbf{x}, \mathbf{y}$.

Hint: based on the equation (4) we can write,

$$f(\mathbf{x}^* + t\mathbf{y}) = f(\mathbf{x}^*) + \langle \bigtriangledown f(\mathbf{x}^*), t\mathbf{y} \rangle + \frac{t^2}{2} \langle \mathbf{y}, \bigtriangledown^2 f(\mathbf{x}^*)\mathbf{y} \rangle + o(t^2)$$

b) Show that if f is twice continuously differentiable, then f is convex if and only if the Hessian $\bigtriangledown^2 f(\mathbf{x})$ is positive semi-definite for all $\mathbf{x} \in \mathbb{R}^d$.

Hint: the proof has two steps. **Step 1.** Assume that f is convex. Then prove that $\bigtriangledown^2 f(\mathbf{x})$ is psd. In this step, you should start with the following,

$$f(\mathbf{x} + t\mathbf{y}) \geq f(\mathbf{x}) + \langle \bigtriangledown f(\mathbf{x}), t\mathbf{y} \rangle$$

For the left hand side, we can use equation (4). **Step 2.** Assume that $\bigtriangledown^2 f(\mathbf{x})$ is psd. Then show that f is convex. To do that, start with equation (5).

c) Consider the function $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T A\mathbf{x} + \mathbf{b}^T\mathbf{x} + c$, where A is a symmetric $d \times d$ matrix. Derive the Hessian of f. Under what conditions on A is f convex? Strictly convex?