CS229 Machine Learning (summer 2020): Problem Set 0
v1.1

Andrew ID: Moroccan Student
Name: [Belcaid Anass]
Collaborators: [Working alone]

July 19, 2020

1 Gradient and Hessians [0 points]

1. Let $f(x) = \frac{1}{2}x^T Ax + b^T x$, where A is a symmetric matrix and $b \in \mathbb{R}^n$ is a vector. What is $\nabla f(x)$?

Solution

$$\nabla f(x) = Ax + b \tag{1}$$

2. Let $f(x) = g(h(x))$, where $g : \mathbb{R} \to \mathbb{R}$ is differentiable and $h : \mathbb{R}^n \to \mathbb{R}$ is differentiable. What is $\nabla f(x)$?

Solution

$$\nabla f(x) = \frac{\partial g(h(x))}{\partial h(x)} \frac{\partial h(x)}{\partial x} = g'(x)\nabla h(x) \tag{2}$$

3.

4. Let $f(x) = \frac{1}{2}x^T Ax + b^T x$, where A is a symmetric matrix and $b \in \mathbb{R}^n$ is a vector. What is $\nabla^2 f(x)$?

Solution

$$\nabla^2(f(x)) = A \tag{3}$$

5. Let $f(x) = g(a^T x)$, where $g : \mathbb{R} \to \mathbb{R}$ is continuously differentiable and $a \in \mathbb{R}^n$ is a

vector. What is $\nabla f(x)$ and $\nabla^2(f(x))$?

Solution

$$\nabla g(x) = \frac{\partial g(x)}{\partial a^T x}\frac{\partial a^T x}{\partial x} = g^{'}(a^T x)a \tag{4}$$

$$\nabla^2(g(x))_{(i,j)} = g^{''}(a^T x)a_i a_j \tag{5}$$

2 Positive definite matrices [0 Points]

- Let $z \in \mathbb{R}^n$. Show that $A = zz^T$ is positive semidefinite.

Solution

Let $x \in \mathbb{R}^n$,

$$\begin{aligned} x^T A x &= x^T z z^T x \\ &= (z^T x)^T (z^T x) \\ &= \|z^T x\| \geq 0 \end{aligned}$$

Which prove that A is **SPD**.

- Let $A = zz^T$, what is the null space of A? What is the rank of A?

Solution

The null space $\mathcal{N}(A)$ is given by:

$$\begin{aligned} \mathcal{N}(A) &= \left\{x \in \mathbb{R}^n \mid (zz^T)x = 0\right\} && (6) \\ &= \left\{x \in \mathbb{R}^n \mid \sum_{i=1}^{n} x_i z_i = 0\right\} && (7) \\ &= z^{\perp} && (8) \\ & && (9) \end{aligned}$$

The rank of A is n - the number of **non null** elements in z.

- Let $A \in \mathbb{R}^{n,n}$ be a semidefinite matrix and $B \in \mathbb{R}^{m,n}$ be an arbitrary matrix. Is BAB^T **PSD**?

Solution

For any $x \in \mathbb{R}^m$, let compute the:

$$x^T BAB^T x = (B^T x)^T A(B^T x) \geq 0 \tag{10}$$

Since A is **PSD**,

3 Eigen vectors and Eigen values

- Suppose $A \in \mathbb{R}^{n,n}$ is diagonalizable, that is, $A = TDT^{-1}$ for an invertible matrix T. Show that $At^{(i)} = \lambda_i t^{(i)}$.

Solution

$$\begin{aligned}
At^{(i)} &= TDT^{-1}t^{(i)} && (11)\\
&= TD(T^T t^{(i)}) && (12)\\
&= TD\begin{pmatrix}\ldots & 1 & \ldots\end{pmatrix}^T && (13)\\
&= T\begin{pmatrix}\ldots & \lambda_i & \ldots\end{pmatrix}^T && (14)\\
&= \lambda_i t^{(i)} && (15)
\end{aligned}$$

- Two simple equations on digitalization.

4 Probability and multivariate Gaussian

Let $X = (X_1, \ldots, X_n)$ is sampled from a *multivariate Gaussian* distribution with mean μ and covariance $\Sigma \in S^n_+$.

- Describe the random variable $Y = X_1 + X_2 + \ldots + X_n$. What is the mean and variance?