EECS 545 – Machine Learning - Homework #1

Mina Jafari

group member: Joshua Kammeraad

1) **Question 1.**

(a)

(i)

True.

$(A^{-1})^\top = (A^\top)^{-1}$

multiply both sides by $((A^\top)^{-1})^{-1}$ and we know $((A^\top)^{-1})^{-1} = A^\top$

$A^\top(A^{-1})^\top = A^\top(A^\top)^{-1}$

$A^\top(A^{-1})^\top = I$

$(A^{-1}A)^\top = I$

$I^\top = I$

$I = I$

(ii)

False.

$A = \begin{bmatrix} 3 & 1 \\ 7 & 3 \end{bmatrix}$, $B = \begin{bmatrix} 1 & -5 \\ 12 & 27 \end{bmatrix}$

$(A+B)^{-1} = \begin{bmatrix} 0.15306122 & 0.02040816 \\ -0.09693878 & 0.02040816 \end{bmatrix}$

$A^{-1} + B^{-1} = \begin{bmatrix} 1.81034483 & -0.44252874 \\ -3.63793103 & 1.51149425 \end{bmatrix}$

(iii)

True.

If A is symmetric, $A = A^\top$. Prove $A^{-1} = (A^{-1})^\top$.

$AA^{-1} = I$

$(AA^{-1})^\top = I^\top = I$

$(A^{-1})^\top A^\top = I$ and $A^\top = A$

$(A^{-1})^\top A = I$

$(A^{-1})^\top AA^{-1} = IA^{-1}$

$(A^{-1})^\top = A^{-1}$

(b)

$XX^\top = (U\Sigma V^\top)(U\Sigma V^\top)^\top = U\Sigma V^\top V\Sigma^\top U^\top = U\Sigma\Sigma^\top U^\top$

$XX^\top = Q\Lambda Q^{-1}$

$\implies U\Sigma\Sigma^\top U^\top = Q\Lambda Q^{-1}$

$\implies Q = U, Q^{-1} = U^\top, \Lambda = \Sigma\Sigma^\top$

The eigen decomposition results in eigenvalues $\lambda_i = \Lambda_{ii} = \sigma_i^2$ for $1 \leq i \leq m$ (for non-zero diagonal elements of Λ) and columns of U are orthonormal and constitute the eigenvectors of $XX^\top$.

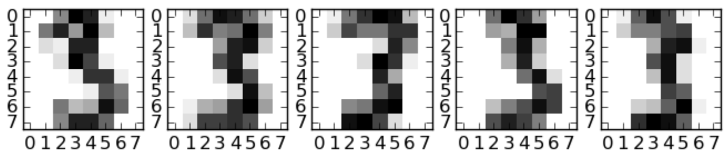
(c)

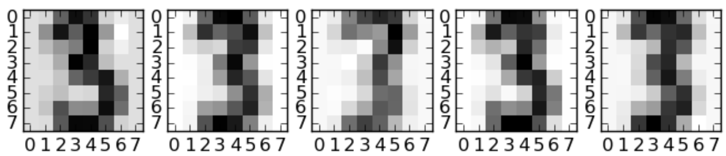
(i)

757.00219118, 158.20556896, 130.31529335

(ii)

$\|A - B\|_F^2 = 68125.6048172$





2) **Question 2.**

(a)

(i)

depends.

If H & D are **independent**, then $\frac{P(H)P(D)}{P(D)}$? $P(H)$, so $P(H = h|D = d) = P(H = h)$.

(ii)

$\geq$

$P(H = h|D = d) = \frac{P(H \cap D)}{P(D)}$

$P(D = d|H = h)P(H = h) = \frac{P(D \cap H)}{P(H)} P(H)$

$\frac{P(H \cap D)}{P(D)}$? $P(D \cap H)$

$\frac{P(D \cap H)}{P(D)}$? $P(D \cap H)$

So, $P(H = h|D = d) \geq P(D = d|H = h)P(H = h)$

(b)

(i)

$$\mathbb{E}[X] = \int_x p(x)X dx$$

$$\mathbb{E}_Y[\mathbb{E}_X[X|Y]] = \int_y p(y) \int_x p(x|y)X dxdy = \int_y \int_x p(y)\frac{p(x,y)}{p(y)} X dxdy$$
$$= \int_x \int_y p(x,y)X dxdy = \int_x p(x)X dx$$

$\implies \mathbb{E}[X] = \mathbb{E}_Y[\mathbb{E}_X[X|Y]]$

(ii)

$\text{var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

$\mathbb{E}_Y[\text{var}_X[X|Y]] + \text{var}_Y[\mathbb{E}_X[X|Y]] =$

$\mathbb{E}_Y[\mathbb{E}_X[X^2|Y] - (\mathbb{E}_X[X|Y])^2] + \mathbb{E}_Y[(\mathbb{E}_X[X|Y])^2] - (\mathbb{E}_Y[(\mathbb{E}_X[X|Y])])^2$

$\mathbb{E}_Y[\mathbb{E}_X[X^2|Y]] - \mathbb{E}_Y[(\mathbb{E}_X[X|Y])^2] + \mathbb{E}_Y[(\mathbb{E}_X[X|Y])^2] - (\mathbb{E}_Y[(\mathbb{E}_X[X|Y])])^2$

middle terms cancel out and based on part **i** of the question:

$\mathbb{E}_Y[\mathbb{E}_X[X^2|Y]] - (\mathbb{E}_Y[(\mathbb{E}_X[X|Y])])^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \text{var}[X]$

3) **Question 3.**

(a)

Prove that if A is PSD, then $\lambda_i \geq 0$:

Assume A is PSD, then $x^\top Ax \geq 0$.

Since A is real and symmetric, we can use the spectral theorem to express A as $A = U\Lambda U^\top$ and $Au_i = \lambda_i u_i$ for all i.

If $x^\top Ax \geq 0$ is true for any vector x, then it should be true for the eigenvectors of A, u_i for all i.

We can write $u_i^\top Au_i \geq 0$ based on $x^\top Ax \geq 0$.

$u_i^\top Au_i \geq 0 \to u_i^\top \lambda_i u_i \geq 0 \to \lambda_i u_i^\top u_i \geq 0 \xrightarrow{\|u_i\|=\sqrt{u_i^\top u_i}} \lambda_i \|u_i\|^2 \geq 0$

Since $\|u_i\| > 0 \implies \lambda_i \geq 0$

Prove that if $\lambda_i \geq 0$, then A is PSD:

Assume $\lambda_i \geq 0$. The spectral theorem holds since A is real and symmetric, then $A = U\Lambda U^\top$ and $Au_i = \lambda_i u_i$ for all i.

$Au_i = \lambda_i u_i \to u_i^\top Au_i = u_i^\top \lambda_i u_i \to u_i^\top Au_i = \lambda_i u_i^\top u_i = \lambda_i$

If $\lambda_i \geq 0 \implies u_i^\top Au_i \geq 0$ for $1 \leq i \leq d$

A is a real and symmetric $d \times d$ matrix. U is an orthogonal $d \times d$ matrix and its columns are eigenvectors of the A, so it spans the whole space of A. Since vectors u_i and x span the same space, we can write $u_i^\top Au_i \geq 0$ into the general form of $x^\top Ax \geq 0$. This property means A is PSD ($A \geq 0$).

(b)

Prove that if A is PD, then $\lambda_i > 0$:

Assume A is PD, then $x^\top Ax > 0$ for all $x \neq 0$.

Since A is real and symmetric, we can use the spectral theorem to express A as $A = U\Lambda U^\top$ and $Au_i = \lambda_i u_i$ for all i.

If $x^\top Ax > 0$ is true for any vector x, then it should be true for the eigenvectors of A, u_i for all i.

We can write $u_i^\top Au_i > 0$ based on $x^\top Ax > 0$.

$u_i^\top Au_i > 0 \to u_i^\top \lambda_i u_i > 0 \to \lambda_i u_i^\top u_i > 0 \xrightarrow{\|u_i\|=\sqrt{u_i^\top u_i}} \lambda_i \|u_i\|^2 > 0$

Since $\|u_i\| > 0 \implies \lambda_i > 0$

Prove that if $\lambda_i > 0$, then A is PD:

Assume $\lambda_i > 0$. The spectral theorem holds since A is real and symmetric, then $A = U\Lambda U^\top$ and $Au_i = \lambda_i u_i$ for all i.

$Au_i = \lambda_i u_i \to u_i^\top Au_i = u_i^\top \lambda_i u_i \to u_i^\top Au_i = \lambda_i u_i^\top u_i = \lambda_i$

If $\lambda_i > 0 \implies u_i^\top Au_i > 0$ for $1 \leq i \leq d$

A is a real and symmetric $d \times d$ matrix. U is an orthogonal $d \times d$ matrix and its columns are eigenvectors of the A, so it spans the whole space of A. Since vectors u_i and x span the same space, we can write $u_i^\top Au_i > 0$ into the general form of $x^\top Ax > 0$. This property means A is PD ($A > 0$).

4) **Question 4.**

$$f(x_i; \lambda) = \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

$$\sum_{i=1}^{n} \log \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

$$\sum_{i=1}^{n} (\log \lambda^{x_i} - \lambda - \log x_i!)$$

$$\frac{df(x_i; \lambda)}{d\lambda} = \sum_{i=1}^{n} (\frac{x_i}{\lambda} - 1) = 0$$

$$\frac{1}{\lambda} \sum_{i=1}^{n} (x_i - n) = 0$$

$$\lambda = \frac{\sum_{i=1}^{n} x_i}{n}$$

5) **Question 5.**

(a)

Strictly convex means $f(\frac{x+y}{2}) < \frac{f(x)}{2} + \frac{f(y)}{2}$.

For a given point z which minimizes the function $f(x)$, $f(z) < f(x \neq z)$.

If the function $f(x)$ has more than one global minimizer (z), then exits a point l in which $f(z) = f(l)$:

$f(\frac{z+l}{2}) < \frac{f(z)}{2} + \frac{f(l)}{2} \rightarrow f(\frac{z+l}{2}) < \frac{f(z)}{2} + \frac{f(z)}{2} \rightarrow f(\frac{z+l}{2}) < f(z)$

There is a point $\frac{z+l}{2}$ such that $f(\frac{z+l}{2}) < f(z)$. So, z is not the global minimizer and $f(x)$ cannot have more than one global minimizer.

(b)

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

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \langle \triangledown f(\mathbf{x}^*), \mathbf{x} - \mathbf{x}^* \rangle + \frac{1}{2}\langle \mathbf{x} - \mathbf{x}^*, \triangledown^2 f(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \rangle + o(\|\mathbf{x} - \mathbf{x}^*\|^2)$$

$$\triangledown f(\mathbf{x}^*) = 0 \implies f(\mathbf{x}) = f(\mathbf{x}^*) + \frac{1}{2}\langle \mathbf{x} - \mathbf{x}^*, \triangledown^2 f(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \rangle + o(\|\mathbf{x} - \mathbf{x}^*\|^2)$$

$$f(\mathbf{x}) - f(\mathbf{x}^*) = \frac{1}{2}\langle \mathbf{x} - \mathbf{x}^*, \triangledown^2 f(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \rangle + o(\|\mathbf{x} - \mathbf{x}^*\|^2)$$

There exists a radius r such that $\|x - x^*\| \leq r \implies f(\mathbf{x}) - f(\mathbf{x}^*) \geq 0$ since x^* is a local minimizer in that radius.

$$\frac{1}{2}\langle \mathbf{x} - \mathbf{x}^*, \triangledown^2 f(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \rangle + o(\|\mathbf{x} - \mathbf{x}^*\|^2) \geq 0$$

$$\frac{1}{2}\langle \|\mathbf{x} - \mathbf{x}^*\| \frac{\mathbf{x} - \mathbf{x}^*}{\|\mathbf{x} - \mathbf{x}^*\|}, \triangledown^2 f(\mathbf{x}^*)(\|\mathbf{x} - \mathbf{x}^*\| \frac{\mathbf{x} - \mathbf{x}^*}{\|\mathbf{x} - \mathbf{x}^*\|}) \rangle + o(\|\mathbf{x} - \mathbf{x}^*\|^2) \geq 0$$

$$\frac{1}{2} \|\mathbf{x} - \mathbf{x}^*\|^2 \langle \frac{\mathbf{x} - \mathbf{x}^*}{\|\mathbf{x} - \mathbf{x}^*\|}, \triangledown^2 f(\mathbf{x}^*)(\frac{\mathbf{x} - \mathbf{x}^*}{\|\mathbf{x} - \mathbf{x}^*\|}) \rangle + o(\|\mathbf{x} - \mathbf{x}^*\|^2) \geq 0$$

$$\frac{\mathbf{x} - \mathbf{x}^*}{\|\mathbf{x} - \mathbf{x}^*\|} \quad \text{and} \quad \triangledown^2 f(\mathbf{x}^*) \quad \text{are constant in a given direction} \implies$$

$$c = \frac{1}{2}\langle \frac{\mathbf{x} - \mathbf{x}^*}{\|\mathbf{x} - \mathbf{x}^*\|}, \triangledown^2 f(\mathbf{x}^*)(\frac{\mathbf{x} - \mathbf{x}^*}{\|\mathbf{x} - \mathbf{x}^*\|}) \rangle, \quad t = \|\mathbf{x} - \mathbf{x}^*\|^2 \quad \text{and}$$

$$\lim_{t \to 0} \frac{o(t)}{ct} = 0 \implies o(t) \text{ goes to zero faster than } ct \implies$$

$$\frac{1}{2}\left\|\mathbf{x}-\mathbf{x}^*\right\|^2\langle\frac{\mathbf{x}-\mathbf{x}^*}{\|\mathbf{x}-\mathbf{x}^*\|},\triangledown^2 f(\mathbf{x}^*)(\frac{\mathbf{x}-\mathbf{x}^*}{\|\mathbf{x}-\mathbf{x}^*\|})\rangle > o(\|\mathbf{x}-\mathbf{x}^*\|^2) \implies$$

if $\frac{1}{2}\left\|\mathbf{x}-\mathbf{x}^*\right\|^2\langle\frac{\mathbf{x}-\mathbf{x}^*}{\|\mathbf{x}-\mathbf{x}^*\|},\triangledown^2 f(\mathbf{x}^*)(\frac{\mathbf{x}-\mathbf{x}^*}{\|\mathbf{x}-\mathbf{x}^*\|})\rangle + o(\|\mathbf{x}-\mathbf{x}^*\|^2) \geq 0 \implies$

$$\frac{1}{2}\left\|\mathbf{x}-\mathbf{x}^*\right\|^2\langle\frac{\mathbf{x}-\mathbf{x}^*}{\|\mathbf{x}-\mathbf{x}^*\|},\triangledown^2 f(\mathbf{x}^*)(\frac{\mathbf{x}-\mathbf{x}^*}{\|\mathbf{x}-\mathbf{x}^*\|})\rangle \geq 0 \implies$$

$$\langle\mathbf{x}-\mathbf{x}^*,\triangledown^2 f(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*)\rangle \geq 0 \implies (\mathbf{x}-\mathbf{x}^*)^\top \triangledown^2 f(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*) \geq 0 \implies \triangledown^2 f(\mathbf{x}^*) \geq 0$$

(c)

Show if $f(\mathbf{x})$ is convex, then $\mathbf{x}^\top \triangledown^2 f(\mathbf{y})\mathbf{x} \geq 0$ for all $\mathbf{y} \in \mathbb{R}^d$.

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

A convex function satisfies $f(\mathbf{x}) \geq f(\mathbf{y}) + \triangledown f(\mathbf{y})\cdot(\mathbf{x}-\mathbf{y})$ for any x, y.

$$f(\mathbf{x}) \geq f(\mathbf{y}) + \langle\triangledown f(\mathbf{y}),\mathbf{x}-\mathbf{y}\rangle \implies$$

$$\frac{1}{2}\langle\mathbf{x}-\mathbf{y},\triangledown^2 f(\mathbf{y})(\mathbf{x}-\mathbf{y})\rangle + o(\|\mathbf{x}-\mathbf{y}\|^2) \geq 0$$

$$\frac{1}{2}\langle\mathbf{x}-\mathbf{y},\triangledown^2 f(\mathbf{y})(\mathbf{x}-\mathbf{y})\rangle + o(\|\mathbf{x}-\mathbf{y}\|^2) \geq 0$$

$$\frac{1}{2}\langle\|\mathbf{x}-\mathbf{y}\|\frac{\mathbf{x}-\mathbf{y}}{\|\mathbf{x}-\mathbf{y}\|},\triangledown^2 f(\mathbf{y})(\|\mathbf{x}-\mathbf{y}\|\frac{\mathbf{x}-\mathbf{y}}{\|\mathbf{x}-\mathbf{y}\|})\rangle + o(\|\mathbf{x}-\mathbf{y}\|^2) \geq 0$$

$$\frac{1}{2}\left\|\mathbf{x}-\mathbf{y}\right\|^2\langle\frac{\mathbf{x}-\mathbf{y}}{\|\mathbf{x}-\mathbf{y}\|},\triangledown^2 f(\mathbf{y})(\frac{\mathbf{x}-\mathbf{y}}{\|\mathbf{x}-\mathbf{y}\|})\rangle + o(\|\mathbf{x}-\mathbf{y}\|^2) \geq 0$$

$\frac{\mathbf{x}-\mathbf{y}}{\|\mathbf{x}-\mathbf{y}\|}$ and $\triangledown^2 f(\mathbf{y})$ are constant in a given direction $\implies$

$c = \frac{1}{2}\langle\frac{\mathbf{x}-\mathbf{y}}{\|\mathbf{x}-\mathbf{y}\|},\triangledown^2 f(\mathbf{y})(\frac{\mathbf{x}-\mathbf{y}}{\|\mathbf{x}-\mathbf{y}\|})\rangle, \quad t = \|\mathbf{x}-\mathbf{y}\|^2$ and

$\lim_{t\to 0}\frac{o(t)}{ct} = 0 \implies o(t)$ goes to zero faster than $ct \implies$

$$\frac{1}{2}\left\|\mathbf{x}-\mathbf{y}\right\|^2\langle\frac{\mathbf{x}-\mathbf{y}}{\|\mathbf{x}-\mathbf{y}\|},\triangledown^2 f(\mathbf{y})(\frac{\mathbf{x}-\mathbf{y}}{\|\mathbf{x}-\mathbf{y}\|})\rangle > o(\|\mathbf{x}-\mathbf{y}\|^2) \implies$$

if $\frac{1}{2}\left\|\mathbf{x}-\mathbf{y}\right\|^2 \langle \frac{\mathbf{x}-\mathbf{y}}{\|\mathbf{x}-\mathbf{y}\|}, \bigtriangledown^2 f(\mathbf{y})(\frac{\mathbf{x}-\mathbf{y}}{\|\mathbf{x}-\mathbf{y}\|})\rangle + o(\|\mathbf{x}-\mathbf{y}\|^2) \geq 0 \implies$

$$\frac{1}{2}\left\|\mathbf{x}-\mathbf{y}\right\|^2 \langle \frac{\mathbf{x}-\mathbf{y}}{\|\mathbf{x}-\mathbf{y}\|}, \bigtriangledown^2 f(\mathbf{y})(\frac{\mathbf{x}-\mathbf{y}}{\|\mathbf{x}-\mathbf{y}\|})\rangle \geq 0 \implies$$

$$\langle \mathbf{x}-\mathbf{y}, \bigtriangledown^2 f(\mathbf{y})(\mathbf{x}-\mathbf{y})\rangle \geq 0 \implies (\mathbf{x}-\mathbf{y})^\top \bigtriangledown^2 f(\mathbf{y})(\mathbf{x}-\mathbf{y}) \geq 0 \implies \bigtriangledown^2 f(\mathbf{y}) \geq 0$$

Show if $x^\top \bigtriangledown^2 f(\mathbf{x})x \geq 0$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, then $f(x)$ is convex.

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

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

if $x^\top \bigtriangledown^2 f(\mathbf{x})x \geq 0 \implies \frac{1}{2}\langle \mathbf{x}-\mathbf{y}, \bigtriangledown^2 f(\mathbf{y}+t(\mathbf{x}-\mathbf{y}))(\mathbf{x}-\mathbf{y})\rangle \geq 0$ for $t \in (0,1) \implies$

$$f(\mathbf{x}) - f(\mathbf{y}) - \langle \bigtriangledown f(\mathbf{y}), \mathbf{x}-\mathbf{y}\rangle \geq 0$$
$$f(\mathbf{x}) \geq f(\mathbf{y}) + \langle \bigtriangledown f(\mathbf{y}), \mathbf{x}-\mathbf{y}\rangle$$
$$f(\mathbf{x}) \geq f(\mathbf{y}) + \bigtriangledown f(\mathbf{y}) \cdot (\mathbf{x}-\mathbf{y})$$

This is the property of a convex function, so $f(x)$ is convex.

(d)

$$f(x) = \frac{1}{2}x^\top Ax + b^\top x + c$$

$$f''(x) = \frac{1}{2}(2A) = A$$

if $y^\top Ay \geq 0$, $f(x)$ is convex for all $x \in \mathbb{R}^d$.
if $y^\top Ay > 0$, $f(x)$ is strictly convex for all $x \in \mathbb{R}^d$.