MACHINE LEARNING
(IN 2064)
EXERCISE SHEET - 1

COLLABORATOR
AKHIL NASSER
03713232

1) $A \in \mathbb{R}^{M \times N}$, $B \in \mathbb{R}^{1 \times M}$, $C \in \mathbb{R}^{N \times P}$,

$D \in \mathbb{R}^{Q \times 1}$, $E \in \mathbb{R}^{N \times N}$, $F \in \mathbb{R}^{1 \times 1}$

2)
$$f(x) = \sum_{i=1}^{N} \sum_{j=1}^{N} x_i x_j M_{ij}$$

$$= \sum_{i=1}^{N} x_i \left(\sum_{j=1}^{N} M_{ij} x_j \right)$$

$$= \sum_{i=1}^{N} x_i (Mx)_i$$

$$\Rightarrow f(x) = x^T M x$$

3) a) The system of linear equations will have unique solution x for any choice of b if

$M = N$ and matrix A is Non-Singular

i.e. $\det(A) \neq 0$

(or)

Multiplying by A^T both sides,

$$A^T A x = A^T b$$

If matrix $A^T A$ is Non-Singular, we will have unique solution for x.

$$\det(A) = \prod_{i=1}^{N} \lambda_i \quad , \quad A \in \mathbb{R}^{N\times N}$$

3.b) As one of the eigen value of $A = 0$, we can say that determinant of $A = 0$.

$\Rightarrow$ A is ~~Non-Si~~ Singular matrix

So, $Ax = b$ will not have Unique solution x for any choice of b.

4)

$$BA = AB = I$$

$$\Rightarrow B = A^{-1}$$

As inverse of matrix A exists, $\det(A) \neq 0$

$$\det(A) = \prod_{i=1}^{n} \lambda_i \neq 0$$

where $\lambda_1, \dots, \lambda_n$ are eigen values of A

$\Rightarrow$ Eigen values of A are non-zero

If $\mu_1, \dots, \mu_n$ are eigen values of B, then $\lambda_i = \frac{1}{\mu_i} \quad \forall i$, where $i = 1, \dots, n$

5) As A is Symmetric matrix, we can write it as $A = U \Lambda U^T$

where U = matrix of eigen vectors of A

& $\Lambda = \text{diag}(\underbrace{\lambda_1, \dots, \lambda_N}_{\Downarrow \text{ Eigen values of } A})$

As A is PSD,

$$x^T A x \geq 0 \qquad x \in \mathbb{R}^N$$
$$A \in \mathbb{R}^{N\times N}$$

$$\Rightarrow x^T U \Lambda U^T x \geq 0$$

$$y^T \Lambda y \geq 0 \quad , \text{where } y = U^T x$$

$$\boxed{\sum_{i=1}^{N} \lambda_i y_i^2 \geq 0}$$

Since y_i^2 is always positive, matrix A should have non-negative eigen values

6) $$B = A^T A \quad \text{(Gram matrix)}$$

In Quadratic form, we get

$$x^T B x = x^T A^T A x \quad , \text{where } x \in \mathbb{R}^N$$

$$= (Ax)^T (Ax)$$

$$= \|Ax\|^2 \geq 0$$

$$\Rightarrow x^T B x \geq 0$$

$\Rightarrow$ Matrix B is positive semi-definite for any choice of A.

7. a) $$f(x) = \frac{1}{2} a x^2 + b x + c$$

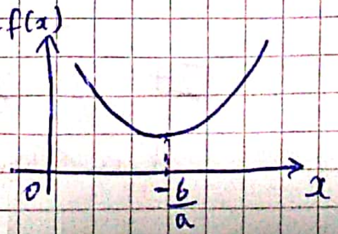
$$f'(x) = ax + b$$

$$f''(x) = a > 0 \quad \text{(To get minimum value of } f(x))$$

For the Optimization problem to have

(i) a Unique solution

$$a > 0$$

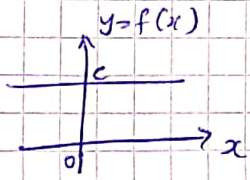


(ii) Infinitely many solutions:

if $a = 0$ and $b = 0$

$\Rightarrow f(x) = c$

We have infinite sol'n of x in this case



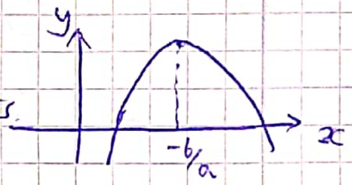
(iii) No Solution:

$a \leq 0$

if $a < 0$, $f(x)$ will look like

Non-convex

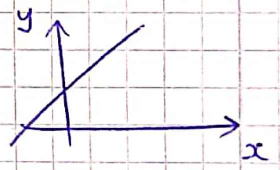
No minima for $f(x)$ exists



Or

if $a = 0$, $f(x)$ ~~will~~ ~~be~~ is a straight line

No minima for $f(x)$ exists



7.b)

$$x^* = \arg\min_{x \in \mathbb{R}} f(x)$$

$$f'(x) = 0$$

$$ax^* + b = 0$$

$$x^* = \frac{-b}{a}$$

8.a)

$$g(x) = \frac{1}{2} x^T A x + b^T x + c$$

It is obvious that Hessian of linear function $b^T x$ and constant c is zero

$\nabla^2 g(x) = ?$

$$\frac{\partial^2 g(x)}{\partial x_k \partial x_\ell} = \frac{\partial}{\partial x_k}\left(\frac{\partial g(x)}{\partial x_\ell}\right) = \frac{\partial}{\partial x_k}\left[\frac{\partial}{2\partial x_\ell}\sum_{i=1}^{N}\sum_{j=1}^{N} A_{ij}x_i x_j + \frac{\partial}{\partial x_\ell}\sum_{i=1}^{N} b_i x_i + 0\right]$$

$$= \frac{\partial}{\partial x_k}\left[\frac{1}{2}\frac{\partial}{\partial x_\ell}\left(\sum_{i\neq\ell}\sum_{j\neq\ell} A_{ij}x_i x_j + \sum_{i\neq\ell} A_{i\ell}x_i x_\ell + \sum_{j\neq\ell} A_{\ell j}x_\ell x_j + A_{\ell\ell}x_\ell^2\right) + b_\ell\right]$$

$$= \frac{\partial}{\partial x_k}\left[\frac{1}{2}\left(\sum_{i\neq\ell} A_{i\ell}x_i + \sum_{j\neq\ell} A_{\ell j}x_j + 2A_{\ell\ell}x_\ell\right)\right] + 0$$

$$= \frac{\partial}{\partial x_k}\left[\frac{1}{2}\left(\sum_{i=1}^{N} A_{i\ell}x_i + \sum_{j=1}^{N} A_{\ell j}x_j\right)\right]$$

$$= \frac{1}{2}\left(A_{k\ell} + A_{\ell k}\right)$$

$$\Rightarrow \nabla^2 g(x) = \frac{1}{2}\left(A + A^T\right)$$

Note: If A is symmetric matrix, $\nabla^2 g(x) = A$

If A is positive definite matrix, the optimization problem will have unique sol'n

8.b) For the existence of solution, matrix A has to be PSD for optimization problem to be well-defined.

The optimization problem will have no solution, if A has a negative eigenvalue

8. c) $\nabla g(x) = 0$

As we already have shown that

$$\frac{\partial g(x)}{\partial x_l} = \frac{1}{2}\left(\sum_{i=1}^{N} A_{il} x_i + \sum_{j=1}^{N} A_{lj} x_j\right) + b_l$$

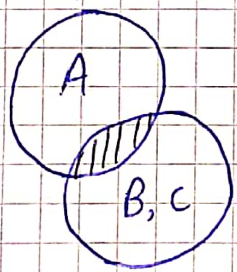
$$\Rightarrow \nabla g(x) = \frac{1}{2}(A + A^T)x^* + b = 0$$

If A is symmetric matrix, we get closed form expression for x^* as

$$Ax^* + b = 0$$

9) $$P(A \mid B, C) = P(A \mid C)$$

Let us use Venn diagrams.
We get the following Venn diagram from the statement above



Now, we can clearly see that

$$\underset{\text{(shaded part)}}{P(A \mid B)} \neq P(A)$$

Thus, $P(A \mid B, C) = P(A \mid C) \Rightarrow P(A \mid B) \neq P(A)$

10) $P(A|B) = P(A)$

$$\frac{P(A \cap B)}{P(B)} = P(A)$$

$$\Rightarrow P(A \cap B) = P(A)\,P(B)$$

we have A and B as independent events

Let us consider rolling of fair die

Let $A = \{2, 4, 6\}$

and $B = \{1, 2, 3, 4\}$

clearly $P(A, B) = P(\{2, 4\}) = \frac{2}{6} = \frac{1}{3}$

$P(A) = \frac{\cancel{3}^{1}}{\cancel{6}_{2}}$ $\quad P(B) = \frac{\cancel{4}^{2}}{\cancel{6}_{3}}$

$$P(A, B) = P(A)\,P(B)$$

To prove $P(A|B, C) = P(A|C)$

Let C be $\{1, 2, 6\}$

$B \cap C = \{1, 2\}$

$$P(A|B, C) = \frac{1}{2}$$

$$P(A|C) = \frac{2}{3}$$

clearly $P(A|B, C) \neq P(A|C)$

Thus, $P(A|B) = P(A) \Rightarrow P(A|B, C) \neq P(A|C)$

11) 1)

$$p(a) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(a,b,c)\,db\,dc$$

2)

$$p(c|a,b) = \frac{p(a,b,c)}{p(a,b)} = \frac{p(a,b,c)}{\int_{-\infty}^{\infty} p(a,b,c)\,dc}$$

3)

$$p(b|c) = \frac{p(b,c)}{p(c)} = \frac{\int_{-\infty}^{\infty} p(a,b,c)\,da}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(a,b,c)\,da\,db}$$

12) Let $d=1$ be event that person has rare disease & $t=1$ be event that test is positive.

$$P(t=1|d=1) = 0.95 \qquad P(d=1) = 0.001$$

$$P(t=1|d=0) = 0.05$$

From Bayes rule,

$$P(d=1|t=1) = \frac{P(t=1|d=1)\,P(d=1)}{P(t=1|d=1)\,P(d=1) + P(t=1|d=0)\,P(d=0)}$$

$$= \frac{0.95 \times 0.001}{0.95\times 0.001 + 0.05 \times 0.999}$$

$\Rightarrow$ Probability that the person has disease on the condition that he obtains a positive result during test $= 0.0187$

13) $E[f(x)] = E[ax + bx^2 + c]$

$$= E[ax] + E[bx^2] + E[c]$$

$$= aE[x] + bE[x^2] + c$$

$$= a\mu + b(\sigma^2 + \mu^2) + c$$

14) a) $E[g(x)] = E[Ax]$ $\quad A \in \mathbb{R}^{N \times N}$

$$= AE[x] \quad x \in \mathbb{R}^N$$

$$= A\mu \quad \mu \in \mathbb{R}^N$$

$$\Sigma = E[(x-\mu)(x-\mu)^T]$$

$$= E[xx^T - x\mu^T - \mu x^T + \mu\mu^T]$$

$$= E[xx^T] - \mu\mu^T - \cancel{\mu\mu^T} + \cancel{\mu\mu^T}$$

$$\Sigma = E[xx^T] - \mu\mu^T \rightarrow (1)$$

b) $E[g(x)g(x)^T] = E[Ax(Ax)^T] = E[Axx^TA^T]$

$$= A\,E[xx^T]\,A^T$$

$$= A(\Sigma + \mu\mu^T)A^T \rightarrow (2)$$

from (1)

c) $E[g(x)^Tg(x)] = E[x^TA^TAx]$

Let $A^TA = B$, $B \in \mathbb{R}^{N \times N}$

Now, $E[x^TBx] = E\left[\sum_{i=1}^{N}\sum_{j=1}^{N} B_{ij}x_ix_j\right]$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N} B_{ij}\,E[x_ix_j]$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} B_{ij} \left(\Sigma_{ij} + \mu_i \mu_j \right) \quad \text{from } 1$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} B_{ij} \Sigma_{ji} + \sum_{i=1}^{N} \sum_{j=1}^{N} B_{ij} \mu_i \mu_j$$

$$= \sum_{i=1}^{N} [B \Sigma]_{i,i} + \mu^T B \mu$$

$$= tr(B \Sigma) + \mu^T B \mu$$

$$E[g(x)^T g(x)] = tr(A^T A \Sigma) + \mu^T A^T A \mu$$

d) $$Cov[g(x)] = cov[Ax]$$

$$= E\left[(Ax - E[Ax])(Ax - E[Ax])^T \right]$$

$$= E\left[(Ax - A\mu)(Ax - A\mu)^T \right]$$

$$= E\left[(Ax - A\mu)(x^T A^T - \mu^T A^T) \right]$$

$$= E\left[A x x^T A^T - A \mu x^T A^T - A x \mu^T A^T + A \mu \mu^T A^T \right]$$

$$= E[A x x^T A^T] - A \mu \mu^T A^T - A \cancel{\mu \mu^T} A^T + A \cancel{\mu \mu^T} A^T$$

$$= A(\Sigma + \mu \mu^T) A^T - A \mu \mu^T A^T$$

$$Cov[g(x)] = A \Sigma A^T$$

Scanned by CamScanner