ASSIGNMENT 1: BASIC MATH

Pranjali Pathre
2019112002.

Question 1: MLE.

1. Given Gaussian distribution.
μ & σ^2 are mean & variance of gaussian distribution.
$\therefore$ we have $N(\mu, \sigma^2)$

Using MLE we have, $\hat{\theta}_{MLE}(x) = \arg\max_{\theta} \log f(x|\theta)$

Log Likelihood $\mathcal{LL} = -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2$

$\arg\max_{\mu} \mathcal{LL}(X|\mu,\sigma^2) :=\frac{\partial \mathcal{LL}}{\partial \mu} = 0$

$$\frac{\partial \mathcal{LL}}{\partial \mu} = \frac{-N}{2}\left(\frac{\partial}{\partial \mu}(\log(2\pi\sigma^2))\right) + \frac{\cancel{2}}{\cancel{2}\sigma^2}\sum_{n=1}^{N}(x_n-\mu).$$

$$= +\frac{1}{\sigma^2}\sum_{n=1}^{N}(x_n-\mu)$$

$$\frac{\partial \mathcal{LL}}{\partial \mu} = 0 \Rightarrow \frac{1}{\sigma^2}\sum_{n=1}^{N}(x_n-\mu) = 0 \Rightarrow \mu = \frac{1}{N}\sum_{n=1}^{N}x_n.$$

$\arg\max_{\sigma^2} \mathcal{LL}(X|\sigma,\mu) := \frac{\partial \mathcal{LL}}{\partial \sigma} = 0.$

$$\frac{\partial \mathcal{LL}}{\partial \sigma} = -\frac{N}{\cancel{2}}\frac{\cancel{2}\cancel{\sigma}(\cancel{2\pi})}{\cancel{2\pi}\sigma^{\cancel{2}}} -$$

$$= \frac{-N}{\cancel{2}}\frac{1}{\sigma} + \frac{2}{2\sigma^3}\sum_{n=1}^{N}(x_n-\mu)^2 = 0$$

$$\Rightarrow \frac{N}{\sigma} = \frac{1}{\sigma^3}\sum_{n=1}^{N}(x_n-\mu)^2 \Rightarrow \sigma^2 = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)^2.$$

2. Let $H_0: Y \sim N(23, \sigma^2)$
$H_1: Y \sim N(33, \sigma^2)$

Let $\mu_1 = 23$ & $\mu_2 = 33$.

$$P(y|0) = \frac{\exp((-(y-\mu_1)^2)/(2\sigma^2))}{\sqrt{2\pi\sigma^2}}, \quad p(y|1) = \frac{\exp((-(y-\mu_2)^2)/(2\sigma^2))}{\sqrt{2\pi\sigma^2}}$$

Likelihood ratio $\mathcal{L}(y) = \frac{p(y|1)}{p(y|0)} = \frac{\exp(-(y-\mu_2)^2/(2\sigma^2))}{\exp(-(y-\mu_1)^2/(2\sigma^2))}$

② $L(y) = \exp\left(-\frac{(y-\mu_2)^2}{2\sigma^2} + \frac{(y-\mu_1)^2}{2\sigma^2}\right)$

$= \exp\left(\frac{-y^2-\mu_2^2+2y\mu_2+y^2+\mu_1^2-2y\mu_1}{2\sigma^2}\right)$

$= \exp\left(\frac{\mu_1^2-\mu_2^2+2y(\mu_2-\mu_1)}{2\sigma^2}\right)$.

∴ For the given problem, $L(y) = \exp($ where $\mu_1 = 23, \mu_2 = 33$.

Log likelihood ratio,

$\log L(y) = \frac{1}{2\sigma^2}(\mu_1^2-\mu_2^2+2y(\mu_2-\mu_1))$.

For MLE rule,

$\log L(y) \underset{H_0}{\overset{H_1}{\gtrless}} 0$

$\Rightarrow \mu_1^2-\mu_2^2+2y(\mu_2-\mu_1) \underset{H_0}{\overset{H_1}{\gtrless}} 0$.

$y \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\mu_2^2-\mu_1^2}{2(\mu_2-\mu_1)} \Rightarrow y \underset{H_0}{\overset{H_1}{\gtrless}} \left(\frac{\mu_2+\mu_1}{2}\right)$.

∴ $y \underset{H_0}{\overset{H_1}{\gtrless}} \left(\frac{23+33}{2}\right)$

$\Rightarrow y \underset{H_0}{\overset{H_1}{\gtrless}} 28$.

3.(i) Now we have $H_0 : N \sim (23, 2\sigma^2)$, $H_1 : N \sim (33, \sigma^2)$

Following same procedure as in part 2.

$1/L(y) = \dfrac{\dfrac{\exp(-(y-\mu_1)^2/2(2\sigma^2))}{\sqrt{2\pi\sigma^2}\,2\sqrt{2}}}{\dfrac{\exp(-(y-\mu_2)^2/2\sigma^2)}{\sqrt{2\pi\sigma^2}}} \Rightarrow$.

$1/L(y) = \frac{1}{\sqrt{2}}\exp\left(-\frac{(y-\mu_1)^2}{4\sigma^2} + \frac{2(y-\mu_2)^2}{4\sigma^2}\right)$

$= \frac{1}{\sqrt{2}}\exp\left(\frac{1}{4\sigma^2}\left(2y^2+2\mu_2^2-4y\mu_2-y^2-\mu_1^2+2y\mu_1\right)\right)$

(3)

$$1/L(y) = \frac{1}{\sqrt{2}} \exp\left(\frac{1}{4\sigma^2}\left(y^2 + 2\mu_2^2 - \mu_1^2 + 2y(\mu_1 - 2\mu_2)\right)\right).$$

$$-\log L(y) = \log\frac{1}{\sqrt{2}} + \frac{1}{4\sigma^2}\left(y^2 + (2\mu_2^2 - \mu_1^2) + 2y(\mu_1 - 2\mu_2)\right).$$

$$-\log L(y) \underset{H_0}{\overset{H_1}{\gtrless}} 0$$

$$\Rightarrow \frac{1}{2\cancel{4}\sigma^2}\left(y^2 + 2y(\mu_1 - 2\mu_2) + (2\mu_2^2 - \mu_1^2)\right) \underset{H_1}{\overset{H_0}{\gtrless}} \frac{1}{\cancel{2}}\log 2$$

$$y^2 + 2y(\mu_1 - 2\mu_2) + (2\mu_2^2 - \mu_1^2) \underset{H_1}{\overset{H_0}{\gtrless}} 2\sigma^2 \log 2.$$

For the given problem, $y^2 + 2y(-43) + (1649) \underset{H_1}{\overset{H_0}{\gtrless}} 2\sigma^2 \log 2.$

$$y^2 - 86y + 1649 \underset{H}{\overset{H_0}{\gtrless}} 2\sigma^2 \log 2.$$

$$\left(y - (43 + 10\sqrt{2})\right)\left(y - (43 - 10\sqrt{2})\right) \underset{H_1}{\overset{H_0}{\gtrless}} 2\sigma^2 \log 2.$$

(ii) We have $H_0 : N \sim (23, \sigma^2)$, $H_1 : N \sim (33, \cancel{\sigma}^2)$.

$$1/L(y) = \frac{\dfrac{\exp(-(y-\mu_1)^2/2\sigma^2)}{\sqrt{2\cancel{\pi}\sigma^2}}}{\dfrac{\exp(-(y-\mu_2)/4\sigma^2)}{\sqrt{2\cancel{\pi}\cdot 2\sigma^2}}} = \frac{2}{\sqrt{2}} \exp\left(-\frac{(y-\mu_1)^2}{2\sigma^2} + \frac{(y-\mu_2)^2}{4\sigma^2}\right)$$

$$= \sqrt{2} \exp\left(\frac{-2(y^2 + \mu_1^2 - 2y\mu_1) + y^2 + \mu_2^2 - 2y\mu_2}{4\sigma^2}\right).$$

$$= \sqrt{2} \exp\left(\frac{-y^2 + \mu_2^2 - 2\mu_1^2 + 4y\mu_1 - 2y\mu_2}{4\sigma^2}\right).$$

$$= \sqrt{2} \exp\left(\frac{-y^2 + (\mu_2^2 - 2\mu_1^2) + 2y(2\mu_1 - \mu_2)}{4\sigma^2}\right).$$

(A)

$$-\log L(y) = \log\sqrt{2} + \frac{-y^2+(\mu_2^2-2\mu_1^2)+2y(2\mu_1-\mu_2)}{4\sigma^2}$$

$$\therefore -\log L(y) \underset{H_0}{\overset{H_1}{\gtrless}} 0.$$

$$\Rightarrow \log\sqrt{2} + \frac{-y^2+(\mu_2^2-2\mu_1^2)+2y(2\mu_1-\mu_2)}{4\sigma^2} \underset{H_1}{\overset{H_0}{\gtrless}} 0.$$

$$\frac{1}{2}\log 2 \underset{\cancel{H_0}H_1}{\overset{H_0}{\gtrless}} \frac{y^2+(2\mu_1^2-\mu_2^2)-2y(2\mu_1-\mu_2)}{4\sigma^2}.$$

$$\frac{1}{2}\log 2 \underset{H_1}{\overset{H_0}{\gtrless}} \frac{y^2 \overset{-31}{\cancel{+1989}} - 2y(13)}{4\sigma^2}$$

$$\frac{1}{\cancel{2}}\log 2 \underset{H_1}{\overset{H_0}{\gtrless}} \frac{y^2-26y\overset{-31}{\cancel{+859}}}{\cancel{2}\sigma^2} \Rightarrow \cancel{(y-(13+\sqrt{7900}))(y-(13-i\sqrt{790}))} \underset{H_0}{\overset{H_1}{\gtrless}} \cancel{2\sigma\log}$$

$$\Rightarrow (y-(13+10\sqrt{2}))(y-(13-10\sqrt{2})) \underset{H_0}{\overset{H_1}{\gtrless}} 2\sigma^2\log 2.$$

$$\cancel{(y-(13+\sqrt{790}i))(y-(13-i\sqrt{790}))} \underset{H_0}{\overset{H_1}{\gtrless}} \cancel{2\sigma^2\log 2}.$$

From above we observe that if $H_0: N\sim(23, 2\sigma^2)$ & $H_1: N\sim(33, \sigma^2)$,

The threshold shifts depending on the value of σ^2. If $\sigma^2>18$, the threshold decreases and otherwise the threshold increases.

Opposite is the case when $H_0: N\sim(23, \sigma^2)$ & $H_1: N\sim(33, 2\sigma^2)$.

Question 2.

Given X, a continuous random variable with PDF:

$$f_X(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}} \quad \forall x \in \mathbb{R}.$$

$Y = X^2$.

1. Find $f_Y(y)$

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y)$$
$$= P(-\sqrt{y} \leq X \leq \sqrt{y})$$
$$= P$$
$$F_Y(y) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$
$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{d}{dy}\left(F_X(\sqrt{y}) - F_X(-\sqrt{y})\right) = f_X(\sqrt{y})\frac{d}{dy}(\sqrt{y}) - f_X(-\sqrt{y})\frac{d}{dy}(-\sqrt{y})$$
$$= \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y})$$

We have $f_X(\sqrt{y}) = \frac{e^{-y/2}}{\sqrt{2\pi}}$ & $f_X(-\sqrt{y}) = \frac{e^{-y/2}}{\sqrt{2\pi}}$.

$$\therefore f_Y(y) = \frac{1}{\cancel{2}\sqrt{y}}\left[\frac{\cancel{2} \cdot e^{-y/2}}{\sqrt{2\pi}}\right] = \frac{1}{\sqrt{2\pi y}} e^{-y/2}.$$

2. $E[X]$.

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx = \int_{-\infty}^{\infty} \frac{x e^{-x^2/2}}{\sqrt{2\pi}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2} dx.$$

Let $u = -\frac{x^2}{2} \Rightarrow \frac{du}{dx} = -x$.

$$\therefore \int x e^{-x^2/2} dx = -\int e^u du = -e^u.$$
$$\therefore \int x e^{-x^2/2} dx = -e^{-\frac{x^2}{2}}.$$
$$\therefore E[X] = \frac{1}{\sqrt{2\pi}}\left[e^{-\frac{x^2}{2}}\right]\Bigg|_{-\infty}^{\infty}$$

6

As the given function is symmetric,

$E[X] = \frac{1}{\sqrt{2\pi}}[0] = 0.$

3. $\sigma^2[X]$

$\sigma^2[X] = E[(X-E(X))^2) = E[X^2] - (E[X])^2.$

$= E[X^2] - 0 = E[X^2] \quad (\because E[X]=0)$

We have $E[X^2] = E[Y]$ as $Y = X^2$.

$\therefore E[Y] = \int_{-\infty}^{\infty} y f_Y(y)\,dy$

$= \int_{-\infty}^{\infty} y \frac{1}{\sqrt{2\pi y}} e^{-y/2}\,dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sqrt{y}\, e^{-y/2}\,dy.$

Let $u = -y/2$.

$\int \sqrt{y}\, e^{-y/2}\,dy = -2^{3/2} \int \sqrt{-u}\, e^{u}\,du$

Let $v = \sqrt{-u}$

$\int \sqrt{-u}\, e^{u}\,du = -2\int v^2 e^{-v^2}\,dv = \sqrt{-u}\,e^{u} - \frac{\sqrt{\pi}\,\mathrm{erf}(\sqrt{-u})}{2}.$

$\therefore -2^{3/2}\int \sqrt{-u}\, e^{u}\,du = \sqrt{2}\sqrt{\pi}\,\mathrm{erf}(\sqrt{-u}) - 2^{3/2}\sqrt{-u}\,e^{u}.$

$= \sqrt{2}\sqrt{\pi}\,\mathrm{erf}\left(\sqrt{\frac{y}{2}}\right) - 2\sqrt{y}\,e^{-y/2}.$

$\therefore \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \sqrt{y}\, e^{-y/2}\,dy = \frac{1}{\sqrt{2\pi}}\left[\sqrt{2\pi}\,\mathrm{erf}\left(\sqrt{\frac{y}{2}}\right) - 2\sqrt{y}\,e^{-y/2}\right]\Big|_{-\infty}^{\infty}$

$= \frac{1}{\sqrt{2\pi}} \cdot \sqrt{2\pi} = 1.$

$\therefore \sigma^2(X) = 1.$

Question 3

1. Let X & Y be the two random Variables.

Given, $\bar{X} = \frac{1+2+3+4+5}{5} = \bar{Y}$

$\bar{X} = \bar{Y} = \frac{15}{5} = 3 = E[X] = E[Y]$

$Var(X) = Var(Y) = \frac{(1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2}{5}$

$= \frac{4+1+0+1+4}{5} = \frac{10}{5} = 2.$

$\therefore Var(X) = Var(Y) = 2.$

$cov(X,Y) = E(XY) - E[X]E[Y].$

$= \frac{1+2^2+3^2+4^2+5^2}{5} - 3^2 = \frac{1+4+9+16+25}{5} - 9$

$= 11 - 9 = 2.$

$\therefore$ Covariance Matrix $= \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$

2. Now we have new data set as $\{[1,1]^T, [2,2]^T, [3,3]^T, [4,4]^T, [5,5]^T, [4,3]^T\}$.

$\therefore E[X] = \frac{15+4}{6} = \frac{19}{6}$

$E[Y] = \frac{15+3}{6} = 3.$

$Var[X] = \frac{(1-19/6)^2 + (2-19/6)^2 + (3-19/6)^2 + (4-19/6)^2 + (5-19/6)^2 + (4-19/6)^2}{6}$

$= \frac{\frac{25}{4} + \frac{9}{4} + \frac{1}{4} + \frac{2}{4} + \frac{9}{4}}{6} = \frac{\cancel{46}\,23}{\cancel{4}\times 6\,2} = \frac{23}{12}.$

Var [Y] =

$Var[X] = \left(1-\frac{19}{6}\right)^2 + \left(2-\frac{19}{6}\right)^2 + \left(3-\frac{19}{6}\right)^2 + 2\left(4-\frac{19}{6}\right)^2 + \left(5-\frac{19}{6}\right)^2 \Big/ 6.$

$= \frac{13^2}{6^2} + \frac{49}{6^2} + \frac{1}{6^2} + \frac{2.25}{6^2} + \frac{121}{6^2} \Big/ 6 = \frac{\cancel{390}\;65}{\cancel{6}.6.6}$

$= \frac{65}{36}.$

$Var[Y] = \frac{(1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2 + (3-3)^2}{6}$

$= \frac{4+1+0+1+4}{6} = \frac{10}{6} = \frac{5}{3}$

$cov(X,Y) = E(XY) - E(X)E(Y)$

$= \frac{1^2+2^2+3^2+4^2+5^2+12}{6} - \frac{19}{6}\cdot 3$

$= \frac{67}{6} - \frac{19}{2} = \frac{10}{6} = \frac{5}{3}.$

$\therefore$ The new covariance matrix is $\begin{bmatrix} 65/36 & 5/3 \\ 5/3 & 5/3 \end{bmatrix}$.

Question 4

1. Given, $f(x,y,z) = 5xy^2z$.

Gradient:

$$\nabla f = \begin{bmatrix} \partial f/\partial x \\ \partial f/\partial y \\ \partial f/\partial z \end{bmatrix} = \begin{bmatrix} 5y^2z \\ 5(2xyz) \\ 5xy^2 \end{bmatrix} = 5\begin{bmatrix} y^2z \\ 2xyz \\ xy^2 \end{bmatrix} = 5y\begin{bmatrix} yz \\ 2xz \\ xy \end{bmatrix}.$$

Hessian:

$$\nabla^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial x \partial z} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} & \frac{\partial^2 f}{\partial y \partial z} \\ \frac{\partial^2 f}{\partial x \partial z} & \frac{\partial^2 f}{\partial y \partial z} & \frac{\partial^2 f}{\partial z^2} \end{bmatrix}$$

$$\frac{\partial f}{\partial x} = 5y^2z, \quad \frac{\partial f}{\partial y} = 10xyz, \quad \frac{\partial f}{\partial z} = 5xy^2.$$

$$\frac{\partial^2 f}{\partial x^2} = 0, \quad \frac{\partial^2 f}{\partial x \partial y} = 10yz, \quad \frac{\partial^2 f}{\partial x \partial z} = 5y^2$$

$$\frac{\partial^2 f}{\partial x \partial y} = 10yz, \quad \frac{\partial^2 f}{\partial y^2} = 10xy, \quad \frac{\partial^2 f}{\partial y \partial z} = 10xy.$$

$$\frac{\partial^2 f}{\partial z^2} = 0$$

$$\therefore \nabla^2 f = \begin{bmatrix} 0 & 10yz & 5y^2 \\ 10yz & 10xz & 10xy \\ 5y^2 & 10xy & 0 \end{bmatrix}.$$

2. given $f(x) = \cos(6x)$ at $x = 0$, $a = 0$.

$$f(x) = f(a) + \frac{f'(a)(x-a)}{1!} + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \frac{f^{iv}(a)}{4!}(x-a)^4 + \frac{f^{v}(a)}{5!}(x-5)^5$$

$$\therefore f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \frac{f^{iv}(0)}{4!}x^4 + \frac{f^{v}(0)}{5!}x^5$$

$f(x) = \cos 6x \quad , f(0) = 1$

$f'(x) = -6\sin 6x \, , \; f'(0) = 0$

$f''(x) = -6^2\cos 6x \, , \; f''(0) = -36$

$f'''(x) = +6^3\sin 6x \, , \; f'''(0) = 0.$

$f^{iv}(x) = 6^4\cos 6x \, , \; f^{iv}(0) = 6^4$

$f^{v}(x) = -6^5\sin 6x \, , \; f^{v}(0) = 0.$

$\therefore f(x) = 1 - \frac{6^2}{2!}x^2 + \frac{6^4 x^4}{4!} \dots$

3. Given, $f(x,y) = 3x - 6y$, Constraint $4x^2 + 2y^2 = 25$

$f_x = 3, \; f_y = -6.$

Let $g(x,y) = 4x^2 + 2y^2$

$g_x = 8x, \; g_y = 4y.$

$\therefore \nabla f = \begin{bmatrix} 3 \\ -6 \end{bmatrix}, \; \nabla g = \begin{bmatrix} 8x \\ 4y \end{bmatrix}.$

$f_x = \lambda g_x \,\&\, f_y = \lambda g_y$

$\Rightarrow \quad 3 = \lambda(8x) \quad , \quad -6 = 4y\lambda$

$\lambda x = 3/8 \quad , \quad \lambda y = -3/2 \, , \; \& \; 4x^2 + 2y^2 = 25.$

$4\left(\frac{3}{8\lambda}\right)^2 + 2\left(\frac{-3}{2\lambda}\right)^2 = 25$

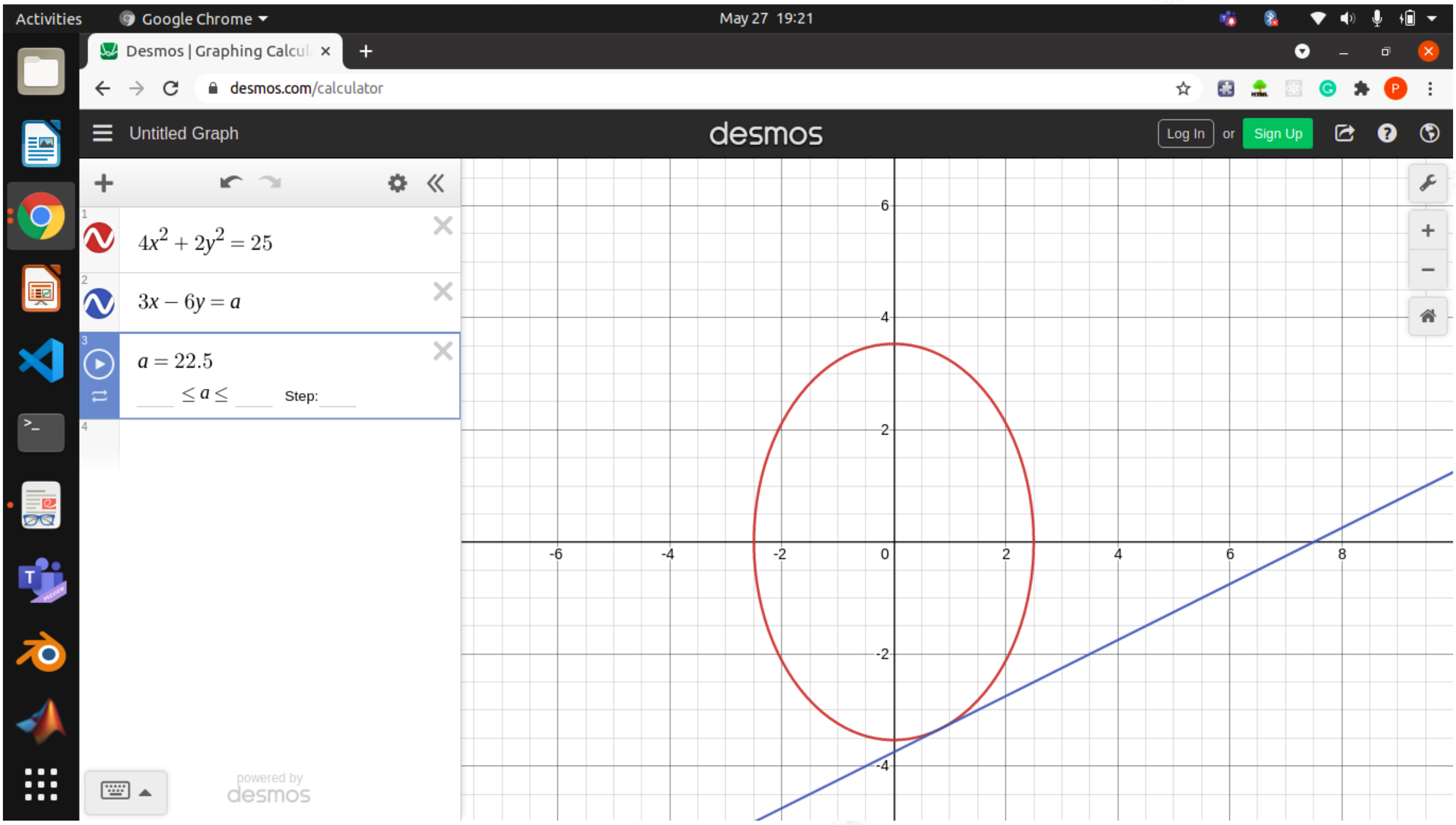
$\Rightarrow \quad 4\frac{36}{64\lambda^2} + \frac{18}{4\lambda^2} = 25 \Rightarrow \lambda^2 25 = \frac{\cancel{36}^{\,9}}{\cancel{64}_{\,16}} + \frac{18\times\cancel{16}}{4\times\cancel{16}} = \frac{81}{16\times 25}$

$\therefore \lambda = \pm\frac{9}{4\times 5} = \pm\frac{9}{20}.$

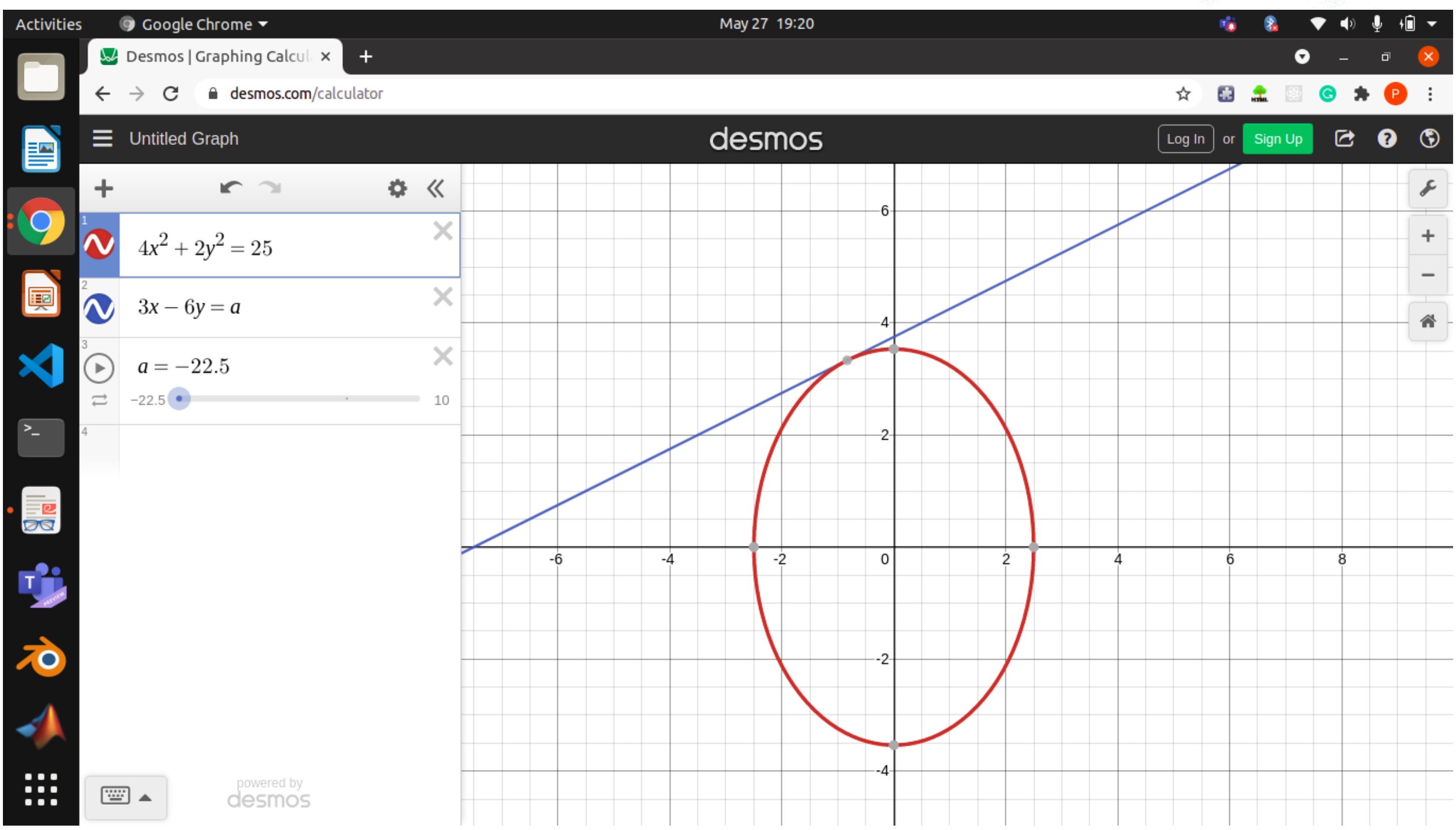
$\therefore x = \pm\frac{\cancel{3}\times \cancel{20}^{10\,5}}{\cancel{8}\times\cancel{9}\,3\;\cancel{4}\,2} = \frac{\pm 5}{6}.$

$y = \pm -\frac{3/2 \times 9}{20} = \pm\frac{\cancel{3}\times\cancel{20}^{10}}{2\times\cancel{9}\,3} = \mp\frac{10}{3}.$

The maximum value of f (x, y) = 3x − 6y



The minimum value of f (x, y) = 3x − 6y



(10) $f(x,y)$ for $\left(\frac{5}{6}, \frac{-10}{3}\right)$ is $f(x,y) = \frac{3\times5}{6\ 2} + \frac{6\times10}{3} = \frac{5}{2} + 20 = 22.5$ (Maximum)

for $\left(\frac{-5}{6}, \frac{10}{3}\right)$ is $f(x,y) = \frac{3\times -5}{6\ 2} - \frac{10}{3}\times 6^2 = \frac{-5}{2} - 20 = -22.5$ (Minimum)

4.. (i)

Let $x_{n+1} \Rightarrow$ Updated value

$x_n \Rightarrow$ Current value.

$\alpha \Rightarrow$ Constant.

$\nabla \Rightarrow$ Gradient.

$f(x_n) \Rightarrow$ given function.

$$x_{n+1} = x_n - \alpha \frac{\partial f(x_n)}{\partial x_n} \quad \text{or}$$

$$x_{n+1} = x_n - \alpha \nabla f(x_n)$$

(ii) $\Rightarrow$ Gradient descent manages to find the minima of function. It evaluated at any point represents the direction of steepest ascent. To minimise the function we can insted follow the negetive of the gradient, and thus go in the direction of steepest descent.

(iii) Let α be the step size.

α controls how big step size we take downhill to reach the minima. If α is very small, then we are taking little baby steps downhill. So multiplying α with gradient tells us the measure of how long ~~the step~~ or the magnitude of the step to be taken to reach the ~~minima~~ downhill.

Depending on the value of α, if it is too large gradient descent can overshoot the minimum, i.e, it may fail to converge or diverge. And if it is too small, gradient descent can be slow.

(iv) Gradient descent cannot tell whether a local or global minima has reached. It finding global or local minima depends on whether ~~we start~~ which initial point we start at.

Question 5

1. Let the entries of matrix S be S_{ij} where it is the element in ith row & jth column.

$$\therefore \quad S = \begin{bmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{bmatrix} \qquad \text{Let } X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

$$\therefore \quad [x_1 \; x_2 \; x_3] \begin{bmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{bmatrix} X = \begin{bmatrix} S_{11}x_1 + S_{21}x_2 + S_{31}x_3 \\ S_{12}x_1 + S_{22}x_2 + S_{32}x_3 \\ S_{13}x_1 + S_{23}x_2 + S_{33}x_3 \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$\therefore \quad x_1(S_{11}x_1 + S_{21}x_2 + S_{31}x_3) + x_2(S_{12}x_1 + S_{22}x_2 + S_{32}x_3) + x_3(S_{13}x_1 + S_{23}x_2 + S_{33}x_3).$$

$$= 4\left((x_1 - 2x_2)^2 + x_3^2\right) + 2(x_1 - 2x_2)x_3$$

$$= \cancel{f}(x$$

$$RHS = 4(x_1^2 + 4x_2^2 - 4x_1x_2 + x_3^2 + 2x_1x_3 - 4x_2x_3)$$

Comparing similar variable terms on both sides,

$$S_{11} = 4, \quad S_{22} = 16, \quad S_{33} = 4$$

$$S_{21} + S_{12} = -16, \quad S_{31} + S_{13} = 8, \quad S_{32} + S_{23} = -16.$$

Therefore we choose S as,

$$S = 4 \begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix}$$

2. From the matrix S, $C_1 + C_2 = -C_3$ & $C_2 = -2C_1$.

$\therefore$ Only two columns of S are linearly independent.

$\therefore$ Rank $= \cancel{2}\; 1$

Finding determinant d, $d = 1(4-4) + 2(-2+2) + 1(+4-4) = 0$.

For finding eigen value, let λ be the eigen value.

$$\therefore \quad |S - \lambda I| = 0$$

$$4\begin{vmatrix} 1-\lambda & -2 & 1 \\ -2 & 4-\lambda & -2 \\ 1 & -2 & 1-\lambda \end{vmatrix} = 0$$

$$\Rightarrow (1-\lambda)((4-\lambda)(1-\lambda)-4) + 2(\cancel{-2}+2\lambda\cancel{+2}) + 1(\cancel{4}-\cancel{4}+\lambda)$$

$$\Rightarrow (1-\lambda)(\cancel{4}+\lambda^2-5\lambda-\cancel{4}) + 4\lambda + \lambda$$

$$= (1-\lambda)(\lambda^2-5\lambda) + 5\lambda$$

$$= \lambda^2 - \cancel{5\lambda} - \lambda^3 + 5\lambda^2 + \cancel{5\lambda}$$

$$= -\lambda^3 + 6\lambda^2 = 0$$

$$\Rightarrow \lambda^3 - 6\lambda^2 = 0 \Rightarrow \lambda^2(\lambda-6) = 0$$

$\therefore$ $\lambda = 0$ & $\lambda = 6$ are the two eigen value of the matrix S.

3. A positive definite matrix is one which satisfies $x^T A x > 0$ for non-zero x.

We have $x^T S x = 4(x_1 - 2x_2 + x_3)^2$.

When $x_1 + x_3 = 2x_2$, $x^T S x = 0$

$\therefore$ $x^T A x \geq 0$ $\therefore$ It is not positive definite matrix. $S \nsucc 0$.

$S \succeq 0$.