Berkay BARLAS
MTK 03734294
Introduction to ML

Problem 1

1.1) $0.5 + 1.6 + 0.0 = 6$

1.2) $\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$

1.3) check $\det(X) \neq 0$

$1.0 - 2.2.8 + 4.8 = 0 \Rightarrow X$ is not invertable

1.4) $Xx = y \quad \underbrace{x = X^{-1}y \mid x = X^{-1}y}_{X \text{ is not invertible}}$

$$\left[\begin{array}{ccc|c} 1 & 2 & 4 & 0 \\ 4 & 1 & 2 & 1 \\ 0 & 1 & 2 & 0 \end{array}\right] \rightarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 4 & 0 & 0 & 1 \\ 0 & 1 & 2 & 0 \end{array}\right]$$

no solution exists

$$\left[\begin{array}{ccc|c} 1 & 2 & 4 & 5 \\ 4 & 1 & 2 & 6 \\ 0 & 1 & 2 & 0 \end{array}\right] \rightarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & 5 \\ 4 & 0 & 0 & 6 \\ 0 & 1 & 2 & 0 \end{array}\right]$$

no solution exists

1.5) $\det(Y - \lambda I) = 0$

$$\left| \begin{bmatrix} 1-\lambda & 2 \\ 2 & 1-\lambda \end{bmatrix} \right| = 0 \rightarrow \begin{array}{l} (1-\lambda)^2 - 4 = 0 \\ \lambda_1 = 1 \quad \lambda_2 = 3 \end{array}$$

$\lambda = 1 \quad (Y - I) \quad \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \begin{array}{l} x_1 = 0 \\ x_2 = 0 \end{array}$

$\lambda = 3 \quad (Y - 3I) \quad \begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad x_1 = x_2 \rightarrow x_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

1-6) $f(x) = (y - \langle z, x \rangle)^2$

$$= 2(y - \langle z, x \rangle) z_1$$

$$\nabla f(x) = 2(y - \langle z, x \rangle) \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$$

1-7) $\nabla f(x) = 0 \rightarrow$ minimizer

$$\nabla f(x) = -2(y - \underbrace{\langle z, x}_{}) z_1 = 0$$

x where $z^T x = 0$ is minimizer

Problem 2

2.1) Sample Mean $\frac{5}{10}$

Sample Variance

$$\left(\frac{1}{2}\right)^2 \cdot 10 \cdot \frac{1}{9} = \frac{5}{18}$$

2.2) In this Order $\rightarrow$ $\frac{1}{\frac{10!}{5!\,5!}} = \frac{1}{252}$

Without Order $\rightarrow p=1$

2.3) $\frac{1}{2}$

The probability func $f(x) = p^x (1-p)^{1-x}$

Likelyhood function $L(\theta) = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i}$

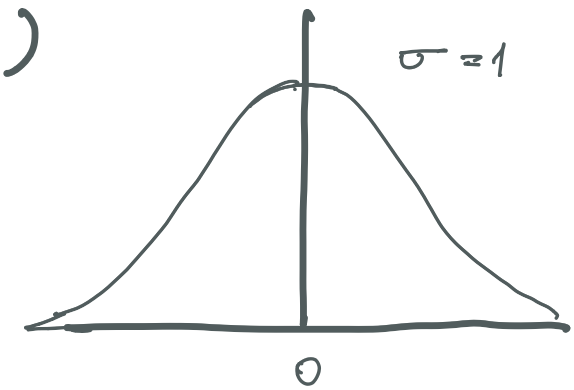
take log $Y \log p + (n-Y) \log(1-p)$ where $Y = \sum_{i=1}^{1} x_i$

the value that maximizes

2.4) If $X \sim N(\mu_x, \sigma_x^2)$
$Y \sim N(\mu_y, \sigma_y^2)$

then
$$X + aY \sim N(\mu_x + a\mu_y,\ \sigma_x^2 + a^2\sigma_y^2)$$

2.5)



Standard normal distribution has 0 mean.
Thus, it is symmetrical according to y-axis.
↳ It's independent of Y

Problem 3

3.1)

(a) $f(x) \overset{(a)}{=} O(g(x))$ and $O(f(x)) \overset{(b)}{=} g(x)$

BOTH

$$f(x) = \log_2(x) = \overbrace{\log_2 e}^{\text{Just Constant}} \cdot \log_e(x) \qquad 1 < \log_2 e < 2$$

(a) for $c = 2$ and $x > 1$ $\quad |f(x)| < c\,g(x)$

(b) for $c = 1$ and $x \geq 1$ $\quad |g(x)| < c\,f(x)$

(b) $O(f(x)) = g(x)$

$$f(x) = e^x = \left(\frac{e}{2}\right)^x \cdot 2^x > g(x) = 2^x$$

for $c = 1$ and $x > 1$ $\quad |g(x)| < c\,f(x)$

(c) $f(x) \underset{(a)}{=} O(g(x))$ and $O(f(x)) \underset{(b)}{=} g(x)$

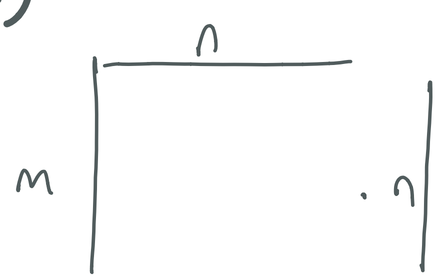
BOTH

$$11x > g(x) = 10x + \log_2(x) > 10x$$

(a) for $c = 1$ and $x > 1$ $\quad |f(x)| < c\,g(x)$

(b) for $c = 12$ and $x > 1$ $\quad |g(x)| < c\,f(x)$

3.2)

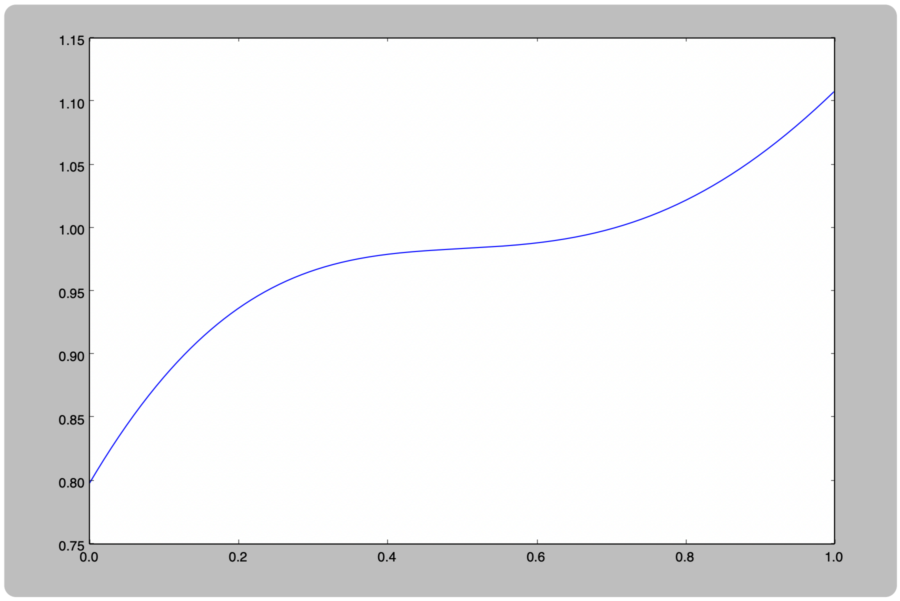


$= m \cdot n$

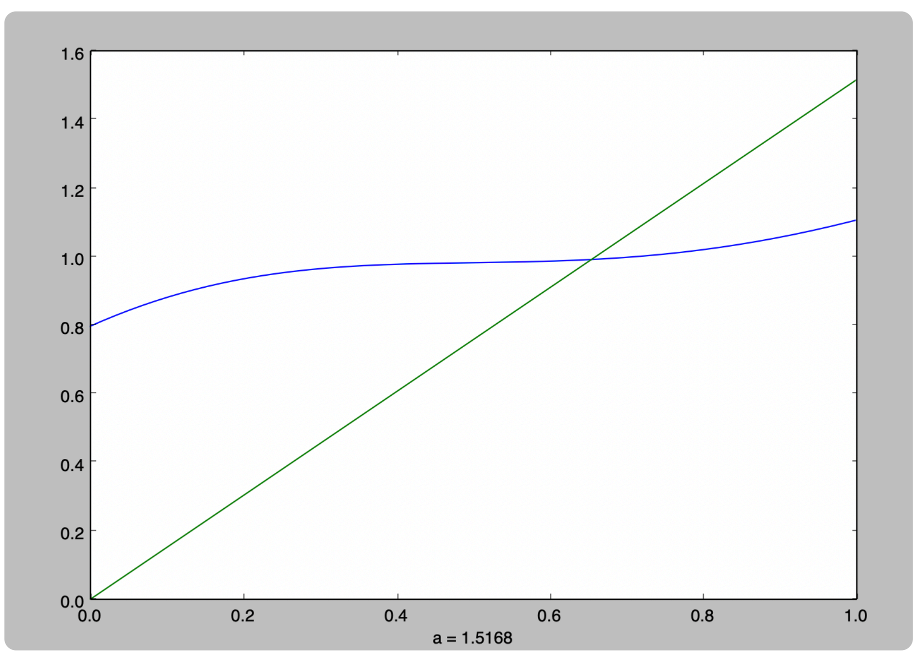
$m \times$ (n multiplication operation)

Problem 4

4.1)



4.2)

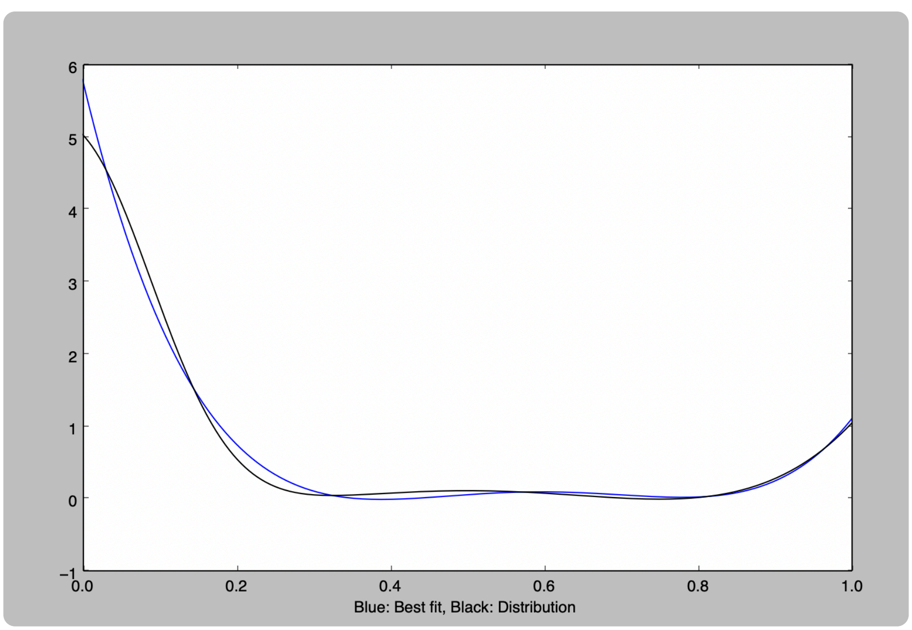


4.3) The solution a is unique when

$$f(a) = \| Xa - y \|_2^2 = \langle Xa - y, Xa - y \rangle$$
$$= \langle Xa, Xa \rangle - 2 \langle Xa, Y \rangle + \langle y, y \rangle$$

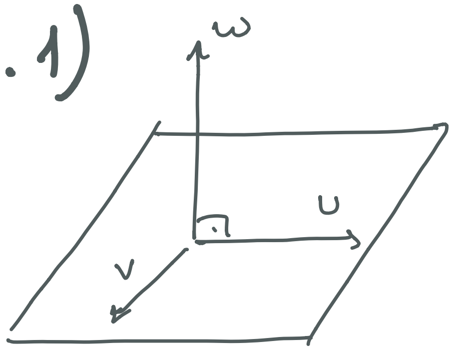
If the rank(X) = rank$(X \mid Y)$ then a is a unique solution where $Y = f(a)$

4.4)



Problem 5

5.1)



$\boxed{\begin{array}{c} w^T u = 0 \\ \text{then } w \text{ and } u \\ \text{orthogonal} \end{array}}$

$$w^T u = -b$$
$$w^T v = -b$$

$\Rightarrow \quad w^T(u-v) = 0$

Thus $u-v$ orthogonal to w

5.2) $\underset{\text{distance}}{d} = \dfrac{|y \cdot w + b|}{\|w\|}$

Assume $x \in R^d$ is a point that satisfies $\langle w, x \rangle + b = 0$

$$d = \|proj_w(y-x)\| = \left\| \frac{(y-x) \cdot w}{w \cdot w} w \right\|$$

$$= \frac{|yw - \overbrace{xw}^{-b}|}{\|w\|}$$

Problem 6

6.1) $Var(aX+b) = a^2\sigma^2$

6.2)

a) According to Central Limit Theorem

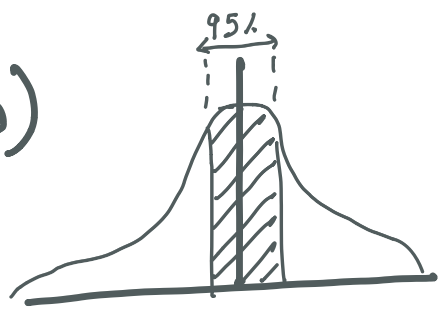
a distribution with mean μ and variance σ^2 will be close to Normal Distribution with large random samples.

$$Y = \sqrt{n}\left(\bar{X} - \frac{n}{2}\right)$$

since $p(\text{heads}) = \frac{1}{2}$ for fair coin

$$E[\bar{X}] = \frac{n}{2}$$

b)



$5000 \cdot \frac{5}{100} = 250$

$5250 > x > 4750$

$$100 \cdot (\bar{X} - 5000) = 4750 \rightarrow \bar{X}_1 = 5047{,}5$$
$$= 5250 \rightarrow \bar{X}_2 = 5052{,}5$$

$$\Rightarrow 5047{,}5 < \bar{X} < 5052{,}5$$

6.3)

a) $0.95 \cdot \frac{0.5}{100} + \frac{1}{10} \frac{99.5}{100} \approx 0.1$

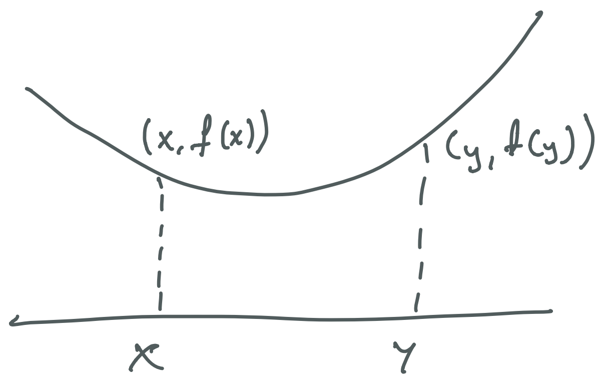
b) $\frac{0.95 \cdot \frac{0.5}{100}}{P(a)} = 0.047$

c) $\frac{\frac{99.5}{100} \cdot \frac{9}{10}}{\frac{0.5}{100} \cdot 0.05 + \frac{99.5}{100} \cdot \frac{9}{10}} \approx 0.999$

d) $0.05 \cdot \frac{0.5}{100} + \frac{1}{10} \cdot \frac{99.5}{100} \approx 0.09975$

Problem 7

7.1)



$\nabla f(x_{min}) = 0$ gives local min since $f(x)$ is convex

(1) $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$

for y point in convex we can choose θ small enough s.t

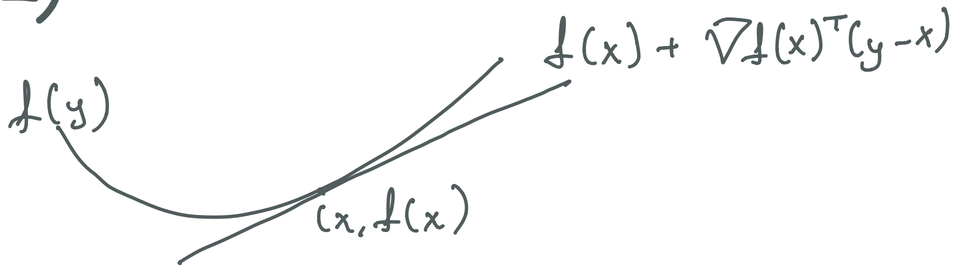
(2) $f(x_{min}) \leq f(x_{min} + \theta(y - x_{min}))$

(1) + (2) $f(x_{min}) \leq \theta f(y) + (1-\theta) f(x_{min})$

$\hookrightarrow$ $f(x_{min}) \leq f(y)$ for any y point in convex which implies x_{min} is global min

7.2)

First Order Condition



$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$$

$$\Downarrow$$

$$f(y + \theta(x-y)) \leq f(y) + \theta(f(x) - f(y))$$

$$f(x) - f(y) \geqslant \frac{f(y+\theta(x-y)) - f(y)}{\theta} \quad \forall \theta \in [0,1]$$

Also assume

$$f(y) \geqslant f(x) + \langle y-x, \nabla f(x) \rangle \quad \forall x, y \text{ in } f(x) \text{ domain}$$

take $z = \theta x + (1-\theta)y$

we have

$$f(x) \geqslant f(z) + \nabla f^T(z)(x-z)$$

$$f(y) \geqslant f(z) + \nabla f^T(z)(y-z)$$

$$\theta f(x) + (1-\theta) f(y) \geqslant f(z) + \nabla f^T(z)(\theta x + (1-\theta)y - z)$$

$$= f(z)$$

$$= f(\theta x + (1-\theta)y) \Rightarrow f \text{ is convex}$$

7.3)

$\theta f(x) + (1-\theta) f(y) \geq f(\theta x + (1-\theta) y)$ is true

since f is convex (part -2) then

$\nabla f(x^*) = 0$ means x^* is the local min

a local min in convex is also a global min.
(part -1)