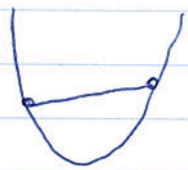
MIDTERM - RE-DO

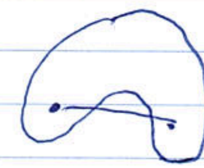
GABE EAPEN

Convex function

Q1]



convex set.



not convex

Optimization Problem

$$\text{minimize } f(x) \quad \text{s.t.} \quad x \in X$$

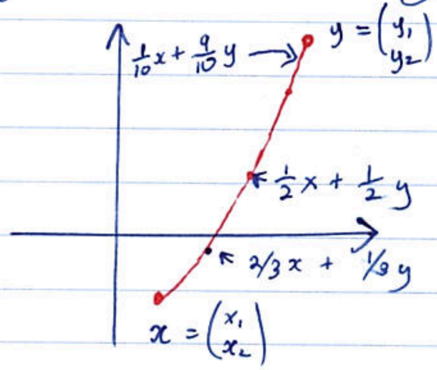
given X is a convex set
given f is convex

Definition:

A set $X \in \mathbb{R}^n$ is convex if $\forall\, x, y \in X$ and $\forall \lambda \in [0,1]$

$$\underbrace{\lambda x + (1-\lambda) y}_{\text{convex combination of } x \text{ and } y} \in X \qquad — (1)$$

What (1) states is that a λ varies between $[0,1]$ a "line segment" is being formed between x and y.



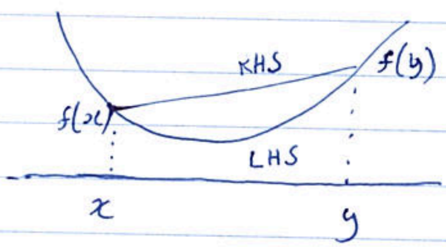
GABE EAPEN

MIDTERM RE-DO.

Definition:
A function $f: \mathbb{R}^n \to \mathbb{R}$ is convex if its domain dom(f) is a convex ~~ses~~ set. and if:

$$\forall x, y \in dom(f) \text{ and } \forall \lambda \in [0,1], \text{ we have}$$

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

convex 

From a geometric standpoint, the line segment connecting $(x, f(x))$ to $(y, f(y))$ should always sit above the graph of the function when convex

So in terms of the question (for proof), the local ~~minimap~~ minimum is also a global minimum

PROOF ATTEMPT:

Let x^* be a local minimum

$\Rightarrow x^* \in X^*$ and $\exists \varepsilon > 0$ s.t $f(x^*) \leq f(x)$ $\forall x \in B(x^*, \varepsilon)$

Suppose for the sake of contradiction, that

$\exists z \in X^*$ with $f(z) < f(x^*)$

But because of given convexity of X, we have

$$\lambda x^* + (1-\lambda) z \in X, \quad \forall \lambda \in [0,1]$$

By convexity of f we have

$$
\begin{aligned}
f(\lambda x^* + (1-\lambda) z) &\leq \lambda f(x^*) + (1-\lambda) f(z) \\
&< \lambda f(x^*) + (1-\lambda) f(x^*) \\
&= f(x^*) \qquad — (2)
\end{aligned}
$$

But as $\lambda \to 1$, $\lambda x^* + (1-\lambda) z \to x^*$ and the previous inequality contradicts local optimality of x^*

Therefore this proves that the set X^* is convex

Q2. a]

$y_i = \beta_0 + \text{noise}$

$$f(\beta) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}(y_i^2 - 2y_i\beta_0 + \beta_0^2)$$

$$\nabla f(\beta) = \frac{1}{n}\sum_{i=1}^{n}\frac{d}{d\beta}\left[y_i^2 - 2y_i\beta_0 + \beta_0^2\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} -2y_i + 2\beta_0$$

$$= \frac{2}{n}\sum_{i=1}^{n} -y_i + \beta_0 \quad \text{--- } (1)$$

Set $\nabla f(\beta_0) = 0$

$$\beta_0 = \frac{2}{n}\sum_{i=1}^{n} y_i \qquad \text{divide by 2}$$

$$= \frac{1}{n}\sum_{i=1}^{n} y_i \qquad \text{without loss of generality}$$

Q2. b]

100 points

sum of 1..100 = $\frac{n(n+1)}{2}$ = 5050

$\bar{x} = \sum_{i=1}^{n} x_i = 50.5$

$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{5050}{100} = 50.5$

$m = \frac{\bar{y}}{\bar{x}} = 1$

$y_i = \beta_0 + m x_i$

Since $y_i = x_i$, $\beta_0 = 0$

c] $\hat{\beta} = (x_i^T x_i)^{-1} x_i^T y_i$

$= (x_i^T x_i)^{-1} x_i^T (x_i \beta^* + e_i)$

$= (x_i^T x_i)^{-1} x_i^T x_i \beta^* + (x_i^T x_i)^{-1} x_i^T e_i$

$= \beta^* + \sum (x_i^T x_i)^{-1} x_i^T e_i$

$= \beta^* + \frac{\sum_i x_i e_i}{\sum_i x_i^2}$

Q3

Let $f_a(x) = \frac{1}{2a} x^2$ — ①

$\nabla f_a(x) = \frac{1}{a} x$ — ②

$$x_{k+1} = x_k - \eta \frac{1}{a} x_k$$

$$= \left(1 - \frac{\eta}{a}\right) x_k$$

$$= \left(\frac{a-\eta}{a}\right) x_k \quad -3$$

Definition:

A function $f: \mathbb{R}^n \to \mathbb{R}$ is called L-Lipschitz if and only if

$$\|\nabla f(x) - \nabla f(y)\|_2 \leq L \|x-y\|_2, \ \forall x, y, \in \mathbb{R}^n$$

<u>Lemma</u> If $f \in C_L$, then $|f(y) - f(x) - \langle \nabla f(x), y-x \rangle| \leq \frac{L}{2} \|y-x\|^2$

From ③

when $a > \eta$, $\left(\frac{a-\eta}{a}\right)$ is positive and increasing

when $a < \eta$ $\left(\frac{a-\eta}{a}\right)$ is negative and decreasing

Q3

PROOF FOR $\eta < a$ convergence.

Let $x^+ = \cancel{x - \eta(\ell}\ x - \nabla f(x)$

Using Lemma,

$$f(x^+) \leq f(x) + \langle \nabla f(x), x^+ - x \rangle + \frac{L}{2} \| x^+ - x \|^2$$

$$= f(x) - \eta \| \nabla f(x) \|^2 + \frac{\eta^2 L}{2} \| \nabla f(x) \|^2$$

$$= f(x) - \eta \left(1 - \frac{\eta}{2} L\right) \| \nabla f(x) \|^2$$

This leads to:

$$\| \nabla f(x) \|^2 \leq \frac{1}{\eta \left(1 - \frac{\eta L}{2}\right)} \left(f(x) - f(x^+) \right)$$

$$\Longrightarrow \sum_{k=1}^{N} \| \nabla f(x^{(k)}) \|^2 \leq \frac{1}{\eta \left(1 - \frac{\eta}{2} L\right)} \left(f(x^{(0)}) - f(x^{(N)}) \right)$$

$$\leq \frac{1}{\eta \left(1 - \frac{\eta}{2} L\right)} \left(f(x^{(0)}) - f^* \right)$$

This implies that:

$$\lim_{k \to \infty} \nabla f(x^{(k)}) = 0$$

If $f(x)$ is convex, $x^{(k)}$ converges to an optimum x^*

~~Which goes ι.~~

P.T.O.

Q3]

Which goes to zero first?

- $f(x^{(k)}) - f^*$
- $\|x^{(k)} - x^*\|$

$\nabla f(x^{(k)})$ goes to 0 as $\frac{c}{\sqrt{k}}$ where c depends on $x^{(0)}$

As $c < 1$,

$$\|x^{(k)} - x^*\| \leq c^{(k)}$$

Q4] a]

Most likely sequence is all heads.

HHH . . . . . . H_n

$$\text{Probability} = \left(\frac{2}{3}\right)^n = \left(\frac{2}{3}\right)^{100}$$

b] Fewer than 50 heads = at most 49 heads

$$x_i = \begin{cases} 1 & \text{Heads} \\ 0 & \text{otherwise} \end{cases}$$

Let X be random variable for one flip.

$$E(X) = \frac{2}{3}(1) + \frac{1}{3}(0) = \frac{2}{3}$$

$$\sigma(X) = \sqrt{Var(X)} = \sqrt{\frac{2}{3}\left(1-\frac{2}{3}\right)^2 + \left(0-\frac{2}{3}\right)^2\frac{1}{3}}$$

$$= \sqrt{\frac{2}{3}\left(\frac{1}{9}\right) + \left(\frac{4}{9}\right)\left(\frac{1}{3}\right)}$$

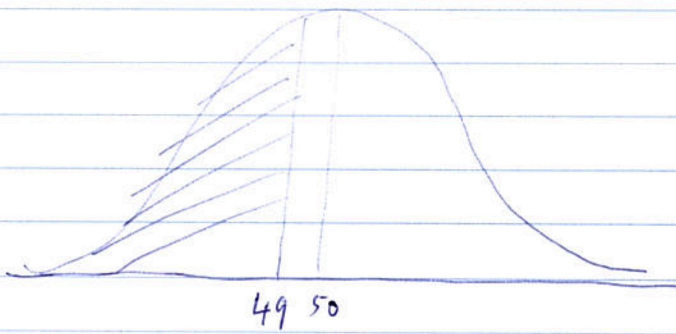
$$= \sqrt{\frac{6}{27}}$$

$$= \frac{1}{3}\sqrt{2}$$

$$\sigma = \frac{\sigma(X)}{n} = \frac{1}{3}\sqrt{2} \times \frac{1}{100} = \frac{\sqrt{2}}{300}$$

Since $n = 100$

$\approx$ Normal distribution with $\mu = \frac{2}{3}$ and $\sigma = \frac{\sqrt{2}}{300}$



$$\boxed{\begin{aligned} E[X_1] &= \frac{2}{3} \\ Var(X_1) &= \frac{6}{27} \end{aligned}}$$ (from p9)

$$P(0 \le S_{100} \le 49) = P\left(0 \le \sum_{i=1}^{100} X_i \le 49\right)$$

$$= P\left(\frac{0 - 100\left(\frac{2}{3}\right)}{\sqrt{\left(\frac{6}{27}\right)(100)}} \le \frac{\sum_{i=1}^{100} X_i - 100\left(\frac{2}{3}\right)}{\sqrt{\left(\frac{6}{27}\right)(100)}} \le \frac{49 - 100\left(\frac{2}{3}\right)}{\sqrt{\left(\frac{6}{27}\right)\cdot(100)}}\right)$$

$$= P(-3.78 \le Z \le 1)$$

$$= \phi(-1) - \phi(-3.78)$$

$$=$$

can be approximated by (where $n = 100$, $k = 49$,

$$\binom{n}{k} p^k (1-p)^{n-k} = \frac{100!}{49!(100-49)!} \left(\frac{2}{3}\right)^k \left(\frac{1}{3}\right)^{n-k}$$

$$= 0.000108$$

Q5] a]

$P(C_1H) = P(C_1H, C_2H) + P(C_1H, C_2T)$

H = 1, Tail = 0 $\quad = \frac{1}{4} + \frac{1}{6} = \frac{3+2}{12} = \frac{5}{12}$

let a_i = outcome head of i^{th} flip

Let Z = number of heads $= \sum_{i=1}^{100} a_i = [a_1 + a_2 + \dots a_{100}]$

$E[Z] = E[a_1 + a_2 + \dots a_{100}] = E[a_1] + E[a_2] + \dots + E[a_{100}]$

$= n(a_i) = 100\left(\frac{5}{12}\right) = \frac{250}{6}$

b] $P(C_2H | C_1H) = \frac{P(C_2H, C_1H)}{P(C_1H)} = \frac{\frac{1}{4}}{5/12} = \frac{1}{4} \times \frac{12}{5} = \frac{3}{5}$

$P(C_2T | C_1H) = \frac{P(C_2T, C_1H)}{P(C_1H)} = \frac{\frac{1}{6}}{5/12} = \frac{1}{6} \times \frac{\overset{2}{12}}{5} = \frac{2}{5}$

$Y = C_2T$
$X = C_2H$

$E[Z] = E[X+Y] = E[X] + E[Y] = \frac{3}{5}(1) + \frac{2}{5}(0)$

$= \underline{\underline{\frac{3}{5}}}$

variance $= p(1-p) = \frac{3}{5}\left(\frac{2}{5}\right) = \frac{6}{25}$

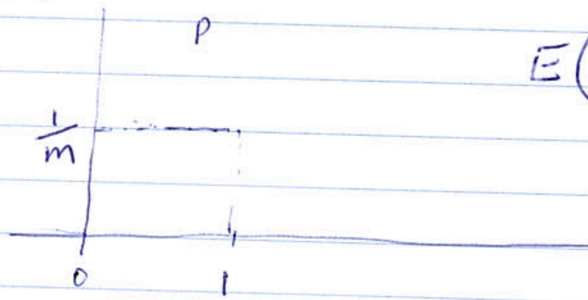
5 c] Uniform distribution for coins in bag

~~PMF~~ is:

$$f(x : 0,1) = \begin{cases} \frac{1}{1-0} & ; \text{ for } 0 \leq x \leq 1 \\ 0 & ; \text{ otherwise} \end{cases}$$

m type of coins

$f(x)$ PMF



$$E(X) = \frac{1}{2}(0+1) = \frac{1}{2}.$$

the probability of heads for a coin with:

$E(X) = \frac{1}{2}$ is $p = \frac{1}{2}$

ASSUME DISCRETE RANDOM VA

$$E[X] = \sum p_i v_i$$
$$= \int_0^1 x\, f_X(x)\, dx$$
$$= \int_0^1 x\, dx = \frac{1}{2}x^2 \Big|_0^1 \frac{1}{2}$$
$$= \frac{1}{2}$$

$\therefore\ p = \frac{1}{2}$

5] d] Let the coin we pick have probability of coming heads as p

Let A = event that the first flip resulted in heads.

Let B = event 2^{nd} flip is heads.

$$P(A) = P(B) = p$$

Find $P(B|A) = \cancel{P(B|A)\,P(A)}\ \frac{P(A|B)\,P(B)}{P(A)}$ (1)

$$\cdot P(A \cup B) = P(A) + P(B) - \underbrace{P(A \cap B)}_{P(A)\,P(B)} = 2p - p^2$$

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

Substituting into (1)

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

$$= \frac{\cancel{P(A)}\,P(B)}{\cancel{P(A)}} = \underline{\underline{p}}$$