(1)(a)(i) let us define a notion of $P(i,n)$

$P(i,n)=$ probability of reaching \$n before losing given he has \$i now.

Let $g(i) = P(i,n)$

so $g(n)=1$, $g(0)=0$.

Now $\forall i \in [1, n-1]$ $\quad g(i) = p \cdot g(i+1) + (1-p) g(i-1)$

$\Rightarrow g(i) - (1-p) g(i-1) = p \cdot g(i+1)$

$\Rightarrow (1-p)(g(i) - g(i-1)) = p \cdot (g(i+1) - g(i))$

$\Rightarrow g(i+1) - g(i) = \left(\frac{1-p}{p}\right)(g(i) - g(i-1))$

$\Rightarrow g(i+1) - g(i) = \beta(g(i) - g(i-1))$

$\Rightarrow g(i+1) - g(i) = \beta^i (g(1) - g(0))$

$\Rightarrow \boxed{g(i+1) - g(i) = \beta^i g(1)}$

put $i=1 \rightarrow g(2) - g(1) = \beta \cdot g(1)$

$i=2 \rightarrow g(3) - g(2) = \beta^2 \cdot g(1)$

$\vdots$

$i=i \rightarrow g(i+1) - g(i) = \beta^i \cdot g(1)$

Sum $\quad g(i+1) = g(1)(1+\beta+\cdots+\beta^i)$

$= g(1)(1+\beta+\cdots+\beta^i) \longrightarrow$ eq (1) now put $i = n-1$

$$\Rightarrow g(n) = g(1)\left(1+\beta+\beta^2+\cdots+\beta^{n-1}\right)$$

$$\Rightarrow g(1) = \frac{g(n)}{1+\beta+\beta^2+\cdots+\beta^{n-1}} = \frac{1}{1+\beta+\beta^2+\cdots+\beta^{n-1}}$$

now we need $g(Y)$ since he has $\$Y$ initially.

so put $i = Y-1$ in Eq (1)

$$\Rightarrow g(Y) = g(1)\left(1+\beta+\cdots+\beta^{Y-1}\right)$$

$$\Rightarrow g(Y) = \frac{1+\beta+\cdots+\beta^{Y-1}}{1+\beta+\cdots+\beta^{n-1}}$$

In our question we need to win extra X dollars so $n = X+Y$

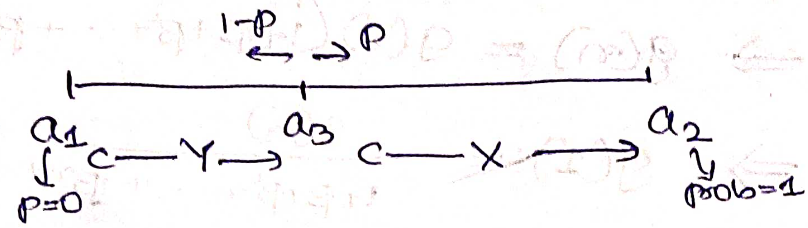
$$\Rightarrow g(Y) = \begin{cases} \dfrac{Y}{X+Y} & \text{if } \cancel{\beta \neq 1}\ \beta = 1 \\[2ex] \dfrac{1-\beta^{Y}}{1-\beta^{X+Y}} & \text{if } \beta \neq 1 \end{cases}$$

~~In our question~~ $\cancel{p = \frac{1}{2}}$ $\Rightarrow \beta = 1$

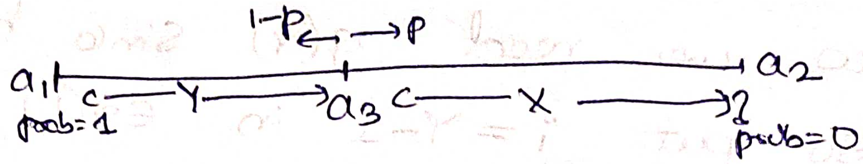
~~so~~ $\cancel{P(Y, X+Y) = \dfrac{Y}{X+Y}}$

(ii) It is similar to above question.

In above question



In our question



now we need to move in opposite direction

so $$g(Y) = \begin{cases} \dfrac{1-\left(\frac{1}{\beta}\right)^{X}}{1-\left(\frac{1}{\beta}\right)^{X+Y}} & \text{if } \beta \neq 1 \\[2ex] \dfrac{Y}{X+Y} & \text{if } \beta = 1 \end{cases}$$

~~We have $\beta = 1$ hence~~

$$\cancel{\boxed{g(Y) = \frac{Y}{X+Y}}}$$

(iii) ~~In this case "no" because probabilities in (i) and (ii) add up to 1.~~
~~But it need not be always true.~~
~~It's happening only because $\beta = 1$ in our case.~~

1 (a) (iii) add probabilities from (i) and (ii)

We get
$$\frac{1-\beta^{Y}}{1-\beta^{X+Y}} + \frac{\beta^{Y}-\beta^{X+Y}}{1-\beta^{X+Y}} = 1$$

(or)

$$\frac{Y}{X+Y} + \frac{X}{X+Y} = 1$$

Hence the probability of playing for ever = 0

(a) (iv) A when $\beta \neq 1$

$$E(\text{gain}) = \left(\frac{1-\beta^{Y}}{1-\beta^{X+Y}}\right)X - \beta^{Y}\left(\frac{1-\beta^{X}}{1-\beta^{X+Y}}\right)Y$$

put $X=1$

$$= \left(\frac{1-\beta^{Y}}{1-\beta^{X+Y}}\right)(Y+1) - Y$$

now
$$\frac{\beta^{1+Y}-1}{\beta^{Y}-1} > \frac{(\beta^{1+Y}-1)+1}{(\beta^{Y}-1)+1}$$

$$> \beta$$

$$\Rightarrow \frac{\beta^{Y}-1}{\beta^{1+Y}-1} < \frac{1}{\beta}$$

$$\Rightarrow \left(\frac{1-\beta^{Y}}{1-\beta^{1+Y}}\right)(Y+1) - Y < \frac{Y+1}{\beta} - Y$$

So $E(\text{gain}) = O\left(\frac{Y+1}{\beta} - Y\right)$

* when $\beta = 1$

$$E(\text{gain}) = \frac{Y}{X+Y}X - \frac{X}{X+Y}Y = 0$$

hence $E(\text{gain}) < \frac{Y+1}{\beta} - Y$

$$\Rightarrow E(\text{gain}) = O\left(\frac{Y+1}{\beta} - Y\right)$$

So in both cases

$$E = O\left((Y+1)/\beta - Y\right)$$

(2) Let us suppose we are betting fraction f, if we have money Z then if we win our money becomes $Z(1+ft)$, and it becomes $Z(1-f)$ if we loose.

Suppose in a single run of n matches we have won w times and lost l times, this w, l are random. So in a single run the amount we would have is

$$Z(1+ft)^{w}(1-f)^{l} \left\{ \text{Here } t = \frac{\text{money gained extra}}{\text{money bet}} \right\}$$

now we need to maximize it in the long run when $n \to \infty$.

→ We know w, l takes random values but as $n \to \infty$ $w \to np$, $l \to nq$

So we need to maximize $(1+ft)^{np}(1-f)^{nq}$

$$\Updownarrow$$

maximizing $(1+ft)^{p}(1-f)^{q}$

$$\Updownarrow$$

maximizing $\log\left((1+ft)^{p}(1-f)^{q}\right)$

$$\Updownarrow$$

$$p\log(1+ft) + q\log(1-f)$$

differentiate w.r.t f

$$\Rightarrow \frac{pt}{1+ft} - \frac{f(1-p)}{1-f} = 0 \Rightarrow f = p - \frac{(1-p)}{t}$$

now in our problem we need to bet $\frac{1}{2} - \frac{1}{2.8} \cdot \frac{1}{2} \cdot \frac{1}{1}$ ≈ 0.14 of our money everytime.

{Note: This is called Kelly Criteria}

Assignment 1

(3) (a) $\operatorname{trace}(XY - YX) = \operatorname{trace}(XY) - \operatorname{trace}(YX)$

$$= 0$$

$$\operatorname{trace}(I_{n\times n}) = n$$

Hence there are no square matrices X, Y such that $XY - YX = I_{n\times n}$.

(b) (i) $\lim_{n\to\infty} \frac{\operatorname{tr}(A_n)}{n}$

$$= \lim_{n\to\infty} \frac{a_{11} + a_{22} + \cdots + a_{nn}}{n}$$

$$\Rightarrow E\left[\lim_{n\to\infty} \frac{\operatorname{tr}(A_n)}{n}\right] = \lim_{n\to\infty} E\left[\frac{a_{11} + a_{22} + \cdots + a_{nn}}{n}\right]$$

$$= \frac{1}{5}\cdot 4 + \frac{4}{5}\cdot 5$$

$$= \frac{24}{5}$$

(ii) $\lim_{n\to\infty} \frac{\log(\det(A_n))}{n}$

$$= \lim_{n\to\infty} \frac{\log(a_{11}) + \log(a_{22}) + \cdots + \log(a_{nn})}{n}$$

$$= \frac{1}{5}\log(4) + \frac{4}{5}\log(5)$$

(3a)

Proof of $trace(XY) = trace(YX)$

$$trace(XY) = (XY)_{11} + (XY)_{22} + \cdots + (XY)_{mm}$$

$$= a_{11}b_{11} + a_{12}b_{21} + \cdots + a_{1k}b_{k1}$$

$$+ a_{21}b_{12} + a_{22}b_{22} + \cdots + a_{2k}b_{k2}$$

$$+ \vdots$$

$$+ a_{m1}b_{1m} + a_{m2}b_{2m} + \cdots + a_{mk}b_{km}$$

Now consider collums as a single entity

$$= (YX)_{11} + (YX)_{22} + \cdots + (YX)_{kk}$$

$$= trace(YX)$$

④ ⓐ $\forall v \in V \sum_{w \in V} T(v,w) = 1$ where V is set of nodes

ⓑ $T(V_1,V_1) = \frac{1}{2}$ $\quad T(V_2,V_1) = \frac{1}{3}$ $\quad T(V_3,V_1) = 0$

$T(V_1,V_2) = \frac{1}{2}$ $\quad T(V_2,V_2) = \frac{1}{3}$ $\quad T(V_3,V_2) = \frac{1}{2}$

$T(V_1,V_3) = 0$ $\quad T(V_2,V_2) = \frac{1}{3}$ $\quad T(V_3,V_3) = \frac{1}{2}$

ⓒ.

we can write $p^{(t+1)} = \begin{bmatrix} x^{t+1} \\ y^{t+1} \\ z^{t+1} \end{bmatrix} = \begin{bmatrix} T(V_1,V_1) & T(V_2,V_1) & T(V_3,V_1) \\ T(V_1,V_2) & T(V_2,V_2) & T(V_3,V_2) \\ T(V_1,V_3) & T(V_2,V_3) & T(V_3,V_3) \end{bmatrix} \begin{bmatrix} x^{(t)} \\ y^{(t)} \\ z^{(t)} \end{bmatrix}$

so $p^{(t+1)} = A p^{(t)}$ where $A =$

so $p^{(t)} = A^{t} \cdot p^{(0)}$

Let A have eigen values $\lambda_1, \lambda_2, \lambda_3$ and their Corresponding eigen vectors x_1, x_2, x_3

here $\lambda_1 = 1$, $\lambda_2 = \frac{1}{2}$, $\lambda_3 = \frac{-1}{6}$ and their Corresponding eigen vectors are $\begin{bmatrix} 1 \\ 3/2 \\ 1 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$

We can write $p^{(0)}$ as $c_1 x_1 + c_2 x_2 + c_3 x_3$ as x_1, x_2, x_3 are linearly independent, so

$$p^{(t)} = A^{t}(c_1 x_1 + c_2 x_2 + c_3 x_3)$$

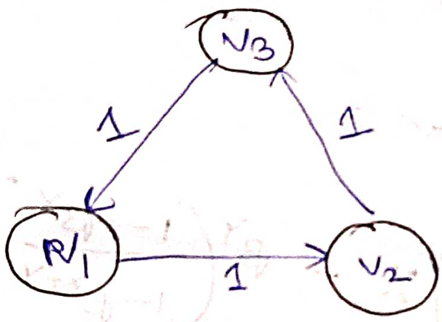
$$= c_1 \lambda_1^{t} x_1 + c_2 \lambda_2^{t} x_2 + c_3 \lambda_3^{t} x_3$$

as $|\lambda_2|, |\lambda_3| < 1$ as $t \to \infty$ $|\lambda_2|^t, |\lambda_3|^t \to 0$

$\Rightarrow \lim_{t \to \infty} p(t) = c_1 x_1$ here $c_1 = \frac{1}{7}$

so $\lim_{t \to \infty} p(t) = \begin{bmatrix} \frac{2}{7} \\ \frac{3}{7} \\ \frac{2}{7} \end{bmatrix}$

(d) this graph doesn't converge for no e_1



5(a) The noise given in the problem is gaussian and by assuming linear regression the correct loss function is $\frac{\sum_{i=1}^{N}(y_i - ax_i - b)^2}{N}$

L2 loss.

5(b) $y = ax + \varepsilon$ $\quad \varepsilon \sim N(0, \sigma^2)$

so $y_j = N(ax_j, \sigma^2)$

Note: Here ε in each training example is independent

$\cancel{P(y_1, y_2 \ldots y_n \mid x_1, x_2 \ldots x_n) = P(y_1 \mid x_1) \cdot P(y_2 \mid x_2)} \ldots$

$=$

A set of i.i.d observations $y = \{y_1, y_2 \ldots y_m\}$ are generated according to a probability model parametrized by a i.e $y_i \sim Pr(y \mid a)$

now $a = \underset{a}{\operatorname{argmax}} \underbrace{\prod_{i=1}^{n} N(y_i \mid ax, \sigma^2)}$

this is Likelihood(a | Data)

Now we need to find Maximum Likelihood Estimate (MLE)

→ Maximizing likelihood is equivalent to Maximizing Log likelihood.

now $a = \underset{a}{\operatorname{argmax}} \sum_{i=1}^{n} \log \left(N(y_i \mid ax, \sigma^2) \right)$

$= \underset{a}{\operatorname{argmax}} \left(k \cdot \sum_{i=1}^{n} -(y_i - ax)^2 \right)$

$= \underset{a}{\operatorname{argmin}} \left(\sum_{i=1}^{n} (ax - y_i)^2 \right)$

let $l = \sum (ax_i - y_i)^2$

$\frac{\partial l}{\partial a} = 0 \Rightarrow \sum 2(ax_i - y_i)(x_i) = 0$

$\Rightarrow \boxed{a = \sum x_i y_i / \sum x_i^2}$

Hence MLE of a $= \sum x_i y_i / \sum x_i^2$

(6) (a) In Simple linear regression model we can take ε as gaussian noise

We ~~But~~ choose ε as $N(0,\sigma^2)$ is preferred because it makes the MLE estimation equivalent to L2 loss.

(b) $$L = \frac{1}{N}\sum_i (a+bx_i-y_i)^2$$

minimizing L:

★ $$\frac{\partial L}{\partial a} = 0 \Rightarrow \frac{2}{N}\sum(a+bx_i-y_i) = 0$$

$$\Rightarrow \frac{2}{N}\left(Na + b\sum x_i - \sum y_i\right) = 0$$

$$\Rightarrow a + b\left(\frac{\sum x_i}{N}\right) - \left(\frac{\sum y_i}{N}\right) = 0 \longrightarrow \text{eq } (1)$$

let $\frac{\sum x_i}{N} = \mu_1$, $\frac{\sum y_i}{N} = \mu_2$

$$\Rightarrow a + b\mu_1 - \mu_2 = 0$$

★ $$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum \frac{2}{N}(a+bx_i-y_i)x_i = 0$$

$$\Rightarrow \frac{2}{N}\left(a\sum x_i + b\sum x_i^2 - \sum x_i y_i\right) = 0$$

$$\Rightarrow a\mu_1 + b\frac{\sum x_i^2}{N} - \frac{\sum x_i y_i}{N} = 0 \longrightarrow \text{eq } (2)$$

We know Corr $r = \frac{\sum x_i y_i - N\mu_1\mu_2}{N\sigma_1\sigma_2}$

where $\sigma_1 = SD(x) = \sqrt{\frac{\sum x_i^2 - N\left(\frac{\sum x_i}{N}\right)^2}{N}} = \sqrt{\frac{\sum x_i^2}{N} - \mu_1^2}$

$\Rightarrow \frac{\sum x_i^2}{N} = \sigma_1^2 + \mu_1^2 \rightarrow$ Eq (5)

Similarly $\frac{\sum y_i^2}{N} = \sigma_2^2 + \mu_2^2$

now $\sum x_i y_i = N r\sigma_1\sigma_2 + N\mu_1\mu_2$

$\frac{\sum x_i y_i}{N} = (r\sigma_1\sigma_2 + \mu_1\mu_2) \rightarrow$ Eq (3)

from Eq (1) $a = \mu_2 - b\mu_1 \rightarrow$ Eq (W)

put Eq (W) in Eq (2), Eq (5) in Eq (2)

$\Rightarrow \mu_1\mu_2 - b\mu_1^2 + b(\sigma_1^2 + \mu_1^2) - (r\sigma_1\sigma_2 + \mu_1\mu_2) = 0$

$\Rightarrow b\sigma_1^2 = r\sigma_1\sigma_2$

$$\Rightarrow \boxed{b = \frac{r\sigma_2}{\sigma_1} = \frac{0.722 \times 10.9}{201} = 0.039}$$

$$\Rightarrow \boxed{a = \mu_2 - \frac{r\sigma_2\mu_1}{\sigma_1} = 57.1 - 0.039 \times (1292.6) = 6.69}$$

ⓒ The one which ~~changes~~ is far ~~the L2~~ loss from our predicted line changes the line most.

- $r=901,\ t=46 \rightarrow 4.3$
- $r=1200,\ t=40 \longrightarrow -13.3$
- $r=1220,\ t=61 \longrightarrow 6.9$
- $r=2400,\ t=80 \longrightarrow -20.1$
- $r=2000,\ t=30 \longrightarrow -581.5$

So $r=2000, t=30$ changes the a value more.

ⓓ ~~P(a,b|Data)~~ ~~P(Data|a,b)~~

$$L(a,b|Data) = P(Data|a,b) \times P(a) \times P(b)$$

$$= \prod_{i=1}^{n} \lambda e^{-\lambda|y_i-a-bx_i|} \times K \times e^{-\frac{(a-\mu_a)^2}{2\sigma_a^2}} \times e^{-\frac{(b-\mu_b)^2}{2\sigma_b^2}}$$

now we need to maximize L which is equivalent to maximizing Log (ML)

$$\Rightarrow \max(-1)\left(\lambda(y_i-a-bx_i) + \frac{(a-\mu_a)^2}{2\sigma_a^2} + \frac{(b-\mu_b)^2}{2\sigma_b^2}\right)$$

$$\Rightarrow \min\left(\lambda(y_i-a-bx_i) + \frac{(a-\mu_a)^2}{2\sigma_a^2} + \frac{(b-\mu_b)^2}{2\sigma_b^2}\right)$$

So the Custom loss function we need to select is

$$\boxed{\frac{1}{N}\sum_{i=1}^{n} \lambda \left|y_i - a - bx_i\right| + \frac{(a-\mu_a)^2}{2\sigma_a^2} + \frac{(b-\mu_b)^2}{2\sigma_b^2}}$$

in our problem it becomes

$$\boxed{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{(a-30)^2}{12} + \frac{b^2}{4} + 2\left(y_i - a - bx_i\right)\right)}$$