1. Probability and Statistics

1. $\mathbb{E}[\text{payout}] = \frac{1}{6}\times 1 + (-\frac{1}{4})\times\frac{5}{6} = -\frac{1}{24} < 0$

So it is not a good bet for us

2 $Pr(\bar{X} \le x) = \int_0^x 4t\,dt \quad 0 \le x \le \frac{1}{2}$

$= 4x^2$

$Pr(\bar{X} \le x) = \int_0^{\frac{1}{2}} 4t\,dt + \int_{\frac{1}{2}}^{x} (-4t+4)\,dt \quad \frac{1}{2} < x \le 1$

$= \frac{1}{2} + [-2x^2 + 4x + \frac{1}{2} - 2]$

$= -2x^2 + 4x - 1$

$\therefore P(\bar{X} \le x) = \begin{cases} 0 & x \le 0 \\ 4x^2 & 0 \le x \le \frac{1}{2} \\ -2x^2+4x-1 & \frac{1}{2} < x \le 1 \\ 1 & x \ge 1 \end{cases}$

3. $Var[\bar{X}] = \mathbb{E}[(\bar{X}-\mu)^2]$

$= \mathbb{E}[\bar{X}^2 - 2\bar{X}\mu + \mu^2]$

$= \mathbb{E}[\bar{X}^2] - 2\mathbb{E}[\bar{X}\mu] + \mathbb{E}[\mu^2]$

$= \mathbb{E}[\bar{X}^2] - 2\mu\,\mathbb{E}[\bar{X}] + \mu^2$

$= \mathbb{E}[\bar{X}^2] - \mu^2 = \mathbb{E}[\bar{X}^2] - (\mathbb{E}[X])^2$

4. $f(x) = ax^2 + bx + c \quad x \sim N(0,1)$

$\mathbb{E}[f(x)] = \mathbb{E}[ax^2] + \mathbb{E}[bx] + \mathbb{E}[c]$

$= a\,\mathbb{E}[x^2] + b\,\mathbb{E}[x] + c$

$= a + c$

2. Proving Stuff

1. $f(x) = \log_e x - x + 1 \quad \forall x > 0$

when $x = 1$ $\quad f(x) = 0 \quad f'(x) = \frac{1}{x} - 1 \quad f''(x) = -\frac{1}{x^2} < 0$

$0 < x < 1 \quad f'(x) > 0 \quad f(x) < f(1) = 0$

$1 < x \quad f'(x) < 0 \quad f(x) < f(1) = 0$

$\therefore f(x) \leq f(1) = 0$ iff $f(x) = 0$ when $x = 1$

2(a) $K(p, q) = \sum_i^k p_i \log\left(\frac{p_i}{q_i}\right)$

$= \sum_i^k p_i \left[-\log\left[\frac{q_i}{p_i}\right]\right]$

$\geq \sum_i^k p_i \left[1 - \frac{q_i}{p_i}\right]$

$= \sum_i^k (p_i - q_i) = \sum_i^k p_i - \sum_i^k q_i = 1 - 1 = 0$

(b) when $p_i/q_i = 1 \quad \forall i \implies p \equiv q$

(c) $p_1 = \frac{1}{2} \quad p_2 = \frac{1}{2} \quad q_1 = \frac{1}{3} \quad q_2 = \frac{2}{3}$

$K(p, q) = \frac{1}{2} \log \frac{3}{2} + \frac{1}{2} \log \frac{3}{4} = \log 3 - \frac{3}{2} \log 2$

$K(q, p) = \frac{1}{3} \log \frac{2}{3} + \frac{2}{3} \log \frac{4}{3} = \frac{5}{3} \log 2 - \log 3$

3. Calculus.

$f(x) = \sigma[\log(5(\max\{x_1, x_2\} \cdot \frac{x_3}{x_4} - (x_5 + x_6))) + \frac{1}{2})$

$= \sigma[\log(5(\max\{5, -1\} \cdot \frac{6}{12} - (7 - 5))) + \frac{1}{2})$

$= \sigma[\log(5(5 \times \frac{1}{2} - 2)) + \frac{1}{2})$

$= \sigma[\log_e(\frac{5}{2}) + \frac{1}{2})$

$\cancel{= \sigma[\log} = \frac{1}{1 + e^{\log_e \frac{2}{5} - \frac{1}{2}}} = \frac{1}{1 + \frac{2}{5} e^{-\frac{1}{2}}} = \frac{5}{5 + 2e^{-\frac{1}{2}}}$

$$x_1 > x_2 \quad \therefore \; f(x) = \sigma\left(\log\left(5\left(x_1 \frac{x_3}{x_4} - (x_5 + x_6)\right)\right) + \frac{1}{2}\right)$$

$$= \sigma\left(\log g(x) + \frac{1}{2}\right)$$

$$\nabla_x f = \sigma(1-\sigma) \cdot \frac{1}{g(x)} \begin{bmatrix} \frac{x_3}{x_4} \\ 0 \\ \frac{x_1}{x_4} \\ -\frac{x_1 x_3}{x_4^2} \\ -1 \\ -1 \end{bmatrix} \Bigg|_{x = x_0}$$

$$= \frac{5}{5+2e^{-\frac{1}{2}}} \cdot \frac{2e^{-\frac{1}{2}}}{5+2e^{-\frac{1}{2}}} \cdot \frac{2}{5} \begin{bmatrix} \frac{1}{2} \\ 0 \\ \frac{5}{12} \\ -\frac{5}{24} \\ -1 \\ -1 \end{bmatrix}$$

4. Softmax Classifier

1. Implementation

$$P_i^j = \frac{e^{z_i^j}}{\sum_k e^{z_i^k}} \qquad L = -\frac{1}{N} \sum_{i=1}^{N} \log P_i^{y_i} = -\frac{1}{N} \sum_{i=1}^{N} \log \frac{e^{z_i^{y_i}}}{\sum_k e^{z_i^k}}$$

$$\frac{\partial L}{\partial z_i^j} = -\frac{1}{N} \frac{\sum_k e^{z_i^k}}{e^{z_i^{y_i}}} \frac{e^{z_i^{y_i}} \sum_k e^{z_i^k} - e^{z_i^{y_i}} e^{z_i^j}}{\left(\sum_k e^{z_i^k}\right)^2} \qquad j = y_i$$

$$= -\frac{1}{N} \frac{\sum_k e^{z_i^k} - e^{z_i^j}}{\sum_k e^{z_i^k}} = \frac{1}{N}\left(P_i^j - 1\right)$$

$$\frac{\partial L}{\partial z_i^j} = -\frac{1}{N} \frac{\sum_k e^{z_i^k}}{e^{z_i^{y_i}}} \frac{-e^{z_i^{y_i}} e^{z_i^j}}{\left(\sum_k e^{z_i^k}\right)^2} = \frac{1}{N} \frac{e^{z_i^j}}{\sum_k e^{z_i^k}} = \frac{1}{N} P_i^j \qquad j \neq y_i$$

since z_i^j is included in $\sum_k$

2. First prove $f(x) = \log \sum_{k=1}^{n} \exp x_k$ is convex

$$\nabla_x f(x) = \frac{1}{\sum_{k=1}^{n} \exp x_k} z \qquad z = \begin{bmatrix} e^{x_1} \\ e^{x_2} \\ \vdots \\ e^{x_n} \end{bmatrix}$$

$$\nabla_x^2 f(x) = \frac{1}{\left(\sum_{k=1}^{n} \exp x_k\right)^2}\left[\left(\sum_{k=1}^{n} \exp x_k\right) \operatorname{diag}(z) - z z^T\right]$$

$$v^T \nabla_x^2 f(x) v = \frac{1}{(1^T z)^2}\left[\left(\sum_{k=1}^{n} z_k\right) \cdot \sum_{k=1}^{n} z_i v_i^2 - \left(\sum_{i=1}^{n} z_i v_i\right)^2\right)$$

$$= \frac{1}{(1^T z)^2}\left(\|a\|_2^2 \|b\|_2^2 - \langle a, b\rangle^2\right) \geq 0$$

$$a = \begin{bmatrix} \sqrt{z_1} \\ \sqrt{z_2} \\ \vdots \\ \sqrt{z_n} \end{bmatrix} \qquad b = \begin{bmatrix} \sqrt{z_1} v_1 \\ \sqrt{z_2} v_2 \\ \vdots \\ \sqrt{z_n} v_n \end{bmatrix}$$

Then suppose there are two w_1, w_2

$w_1 x + b_1 = z_1 \qquad w_2 x + b_2 = z_2$

$$\frac{1}{2}(L(w_1) + L(w_2)) = \frac{1}{2}\left(-z_1^y + \log \sum_k e^{z_1^k} - z_2^y + \log \sum_k e^{z_2^k}\right)$$

$$= \frac{1}{2}\left(-z_1^y - z_2^y + \log \sum_k e^{z_1^k} + \log \sum_k e^{z_2^k}\right)$$

$$\geq \frac{1}{2}\left(-z_1^y - z_2^y + 2\log \sum_k e^{\frac{z_1^k + z_2^k}{2}}\right) = -\left(\frac{z_1^y + z_2^y}{2}\right) + \log \sum_k e^{\frac{z_1^k + z_2^k}{2}}$$

$$= L\left(\frac{w_1 + w_2}{2}\right)$$

$\therefore L$ is convex w.r.t. w