	Predicted NO	Predicted Yes	
Actual NO	50	10	60
Actual Yes	5	100	105
	55	110	

$$Acc = \frac{150}{165} = \frac{10}{11} = 90.909090\%$$

$$Recall = \frac{100}{105} = 95.23\% = \frac{20}{21}$$

$$Precision = \frac{100}{110} = \frac{10}{11} = 90.909090\%$$

$$F1 = \frac{2 \times \frac{20}{21} \times \frac{10}{11}}{\frac{20}{21} + \frac{10}{11}} = \frac{400}{20\times 11 + 10\times 21} = \frac{400}{430}$$

$$= \frac{40}{43} = 0.9302 = 93.02\%$$

```
      0.9302
43 ) 400
    -387
     130
    -129
     100
     -86
      14
```

$$Precision = \frac{\text{TP predicted}}{\text{Total Positive Predicted}}$$

$$Recall = \frac{\text{TP predicted}}{\text{Total Actual Positive}}$$

Type I – False positive

Type II – False Negative

- Maximize Recall when false Negative carries higher cost
  - Screening for terminal disease | Don't want miss anyone
  - Earthquake prediction | Don't want to miss prediction when there is an earthquake
- Maximize Precision when false positive carries higher cost
  - Automatic bombing a target from a drone | should not hurt civillians
  - Giving access to a secure installation | No access to unauthorized personnel

Detection Cost : $C_{FP} \times FP + C_{FN} \times FN$

$P(A) = 0.4$

$P(B) = 0.6$

$P(D|A) = 0.1$

$P(D|B) = 0.05$

$$P(A|D) = \frac{P(D|A)\cdot P(A)}{P(D|A)\cdot P(A) + P(D|B)\cdot P(B)} = \frac{0.1\times 0.4}{0.1\times 0.4 + 0.05\times 0.6} = \frac{0.04}{0.04+0.03}$$

$$= \frac{4}{7} = 0.5714\%$$

$E_1, E_2, \ldots, E_n$ form a partition of a sample space

Then

$$P(E_i|F) = \frac{P(F/E_i)\times P(E_i)}{\sum_{r=1}^{n} P(F/E_r)\times P(E_r)}$$

outlook = sunny, Humidity = High, wind = strong

$P(\text{sunny}|\text{Yes}) = \frac{3}{7}$

$P(\text{high}|\text{Yes}) = \frac{5}{7}$

$P(\text{strong}/\text{Yes}) = \frac{5}{7}$

$P(\text{Yes}) = \frac{7}{14} = 1/2$

$P(\text{No}) = \frac{1}{2}$

$P(\text{sunny}/\text{No}) = \frac{2}{7}$

$P(\text{high}/\text{No}) = \frac{3}{7}$

$P(\text{strong}/\text{No}) = \frac{2}{7}$

$$\begin{array}{r} 5 \\ 87 \\ \times 8 \\ \hline 696 \end{array} \qquad \begin{array}{r} 4 \\ 87 \\ \times 6 \\ \hline 522 \end{array}$$

$$P(\text{Yes}/\text{sunny, high, strong}) = \frac{P(\text{sunny, high, strong}|\text{Yes})\times P(\text{Yes})}{P(s,h,s/\text{Yes}) + P(s,h,s/\text{No})P(\text{No})}$$

$$= \frac{\left(\frac{3}{7}\right)\left(\frac{5}{7}\right)\left(\frac{5}{7}\right)\times\frac{1}{2}}{\left(\frac{3}{7}\right)\left(\frac{5}{7}\right)\left(\frac{5}{7}\right)\times\frac{1}{2} + \frac{2}{7}\left(\frac{3}{7}\right)\left(\frac{2}{7}\right)\left(\frac{1}{2}\right)}$$

$$\Rightarrow \frac{75}{75+12} = \frac{75}{87}$$

$$\begin{array}{r|l} & 0.862 \\ \hline 87 & 750 \\ & -696 \\ \hline & 540 \\ & -522 \\ \hline & 180 \end{array}$$

86

$$P(x_k \mid c=i) = N(\mu_{ik}, \sigma_k^2)$$

$$N(\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$P(c=1 / x_1, \dots, x_d) > P(c=2 / x_1, \dots, x_d)$$

$$P(x_1, \dots, x_d / c=1)\, P(c=1) > P(x_1, \dots, x_d / c=2)\, P(c=2)$$

$$\frac{P(x_1 / c=1) \cdots P(x_d / c=1)}{P(x_1 / c=2) \cdots P(x_d / c=2)} > \frac{P(c=1)}{P(c=2)}$$

$$\frac{\cancel{\frac{1}{\sqrt{2\pi}\,\sigma_1}} e^{-\frac{1}{2}\left(\frac{x-\mu_{1}}{\sigma_1}\right)^2} \cdots \cancel{\frac{1}{\sqrt{2\pi}\,\sigma_d}} e^{-\frac{1}{2}\left(\frac{x-\mu_{1d}}{\sigma_d}\right)^2}}{e^{-\frac{1}{2}}}$$

$$\exp\left(-\frac{1}{2}\left[\left(\frac{x-\mu_{11}}{\sigma_1}\right)^2 - \left(\frac{x-\mu_{21}}{\sigma_1}\right)^2 + x\right.\right.$$

$$O(Nd) + O(\log(N) + \log(N-1) + \cdots + \log(N-K))$$

$$(N)(N-1)(\dots)(N-K)(N-K)!$$

$$O\left(\log\left(\frac{N \cdot (N-1) \cdots (N-K) \cancel{!}}{(N-K)!}\right)\right) \qquad \frac{N!}{(N-K)!}$$

$$O(\log(N!) - \log((N-K)!)) \qquad \log N + \log N-1 + \cdots \log N-K$$

$$N \log N - (N-K)\log(N-K) \qquad \log K(N-K)$$

$$b_1 + b_2 + b_n$$

$$\left(\left(\frac{a_1}{\sqrt{b_1}}\right)^2 + \cdots + \left(\frac{a_1}{\sqrt{b_n}}\right)^2\right) \cdot \left((\sqrt{b_1})^2 + \cdots (\sqrt{b_n})^2\right)$$

$$\geq (a_1 + \cdots a_n)^2 \Rightarrow \left(\frac{a_1^2}{b_1} + \cdots \frac{a_n^2}{b_n}\right) \geq \frac{(a_1 + \cdots a_n)^2}{b_1 + \cdots b_n}$$

$f(\vec{b}_1) = \vec{b}_1^T S \vec{b}_1$ $\quad b_1 = \begin{bmatrix} x \\ y \end{bmatrix}_{2\times 1}$ $\quad S = \begin{matrix} a & b \\ c & d \end{matrix}$

$b_1^T b = 1$

$1\times 2 \quad 2\times 2 \quad 2\times 1$

$f(x,y) = [x\ y]\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$

$= [x\ y]\begin{bmatrix} ax+by \\ cx+dy \end{bmatrix} = ax^2+bxy+cxy+dy^2$

$= ax^2+(b+c)xy+dy^2$

$= \begin{bmatrix} \lambda & b \\ -b & \lambda \end{bmatrix}$ $\quad a+ib \rightarrow \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$

$\boxed{\lambda - bi}$

$\nabla f(x,y) = \begin{bmatrix} 2ax+(b+c)y \\ 2dy+(b+c)x \end{bmatrix} = \lambda\begin{bmatrix} 2x \\ 2y \end{bmatrix}$ $\quad \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \lambda\begin{bmatrix} x \\ y \end{bmatrix}$ $\quad \begin{bmatrix} ax+by \\ cx+dy \end{bmatrix}$

$= \begin{bmatrix} 2a & b+c \\ b+c & 2d \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = 2\lambda\begin{bmatrix} x \\ y \end{bmatrix}$

$= \begin{bmatrix} 2a-2\lambda & b+c \\ b+c & 2d-2\lambda \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = 0.$

$a = \lambda \qquad \underline{b+c=0}$

$d = \lambda$

$z_{1n}^2 = z_{1n}^T z_{1n} \Rightarrow z_{1n} z_{1n}^T$

$b_1^T x_n x_n^T b_1$

$z_{1n} = b_1^T x_n$

$V_1 = \frac{1}{N}\sum_{n=1}^{N} z_{1n}^2$

$V_1 = \frac{1}{N}\sum_{n=1}^{N}(b_1^T x_n x_n^T b_1)$

$= b_1^T \left\{\frac{1}{N}\sum_{n=1}^{N} x_n x_n^T\right\} b_1$

$= b_1^T S b_1$

A is symmetric

$\vec{a} = a_1\vec{x}_1 + a_2\vec{x}_2 + \dots + a_n\vec{x}_n$

↳ arbitrary vector

$\vec{x}_i^T\vec{x}_i = 1 \ \& \ \vec{x}_i^T\vec{x}_j = 0.$

$a_i = \vec{x}_i^T\vec{a}$

$\vec{a} = \sum_{i=1}^{n}(\vec{x}_i^T\vec{a})\vec{x}_i$

$\vec{a} = \left(\sum \vec{x}_i x_i^T\right)\vec{a}$

$\boxed{\sum x_i x_i^T = 0}$

$A\vec{a} = A\left(\sum x_i x_i^T\right)\vec{a}$

$= \left(\sum \lambda_i x_i x_i^T\right)\vec{a}$

$\boxed{A = \lambda_i x_i x_i^T}$

$B = [\vec{b}_1\ \vec{b}_2 \dots \vec{b}_m]$

$\vec{z} = B^T\vec{x}_n \Rightarrow \vec{z}_n = \begin{bmatrix} z_{1n} \\ z_{2n} \\ \vdots \\ z_{mn} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}_{m\times d}\begin{bmatrix} x_{1n} \\ x_{2n} \\ \vdots \\ x_{dn} \end{bmatrix}_{d\times 1}$

$\downarrow$

$z_{1n} = b_1^T\vec{x}_n$

$= \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1d} \\ \vdots & & & \\ b_{m1} & \dots & & b_{md} \end{bmatrix}\begin{bmatrix} x_{1n} \\ \vdots \\ x_{dn} \end{bmatrix} = \begin{bmatrix} b_{11}x_{1n} + b_{12}x_{2n} \dots \\ \vdots \\ b_{m1}x_{1n} + b_{m2}x_{2n} + \dots \end{bmatrix}$

$z_{kn} = b_k^T x_n$

or

$z_{1n} = b_1^T x_n$

$z_{kn} = \begin{bmatrix} b_{k1} \\ \vdots \\ b_{kd} \end{bmatrix} \cdot \begin{bmatrix} x_{1n} \\ \vdots \\ x_{dn} \end{bmatrix}$

$= b_k^T x_n$

$z_{kn} = \sum_{i=1}^{d} b_{ki} x_{in}$

$\vec{f}(\vec{x}) = A\vec{x}$, $\vec{f}(\vec{x}) \in \mathbb{R}^m$, $A \in \mathbb{R}^{M \times N}$, $x \in \mathbb{R}^N$

$f : \mathbb{R}^N \to \mathbb{R}^M$

$$\therefore \vec{f}(\vec{x}) = \begin{bmatrix} f_1(\vec{x}) \\ f_2(\vec{x}) \\ \vdots \\ f_m(\vec{x}) \end{bmatrix}$$

let $\vec{a} = m \times 1$
$\vec{b}^T = 1 \times n$
$\vec{b} = n \times 1$

$\underset{1\times m}{a^T} \times \underset{n \times 1}{b}$

$m \times n$

$$\frac{\vec{f}(\vec{x})}{\partial x_i} = \begin{bmatrix} \lim_{h\to 0} f_1(x_1, \dots, x_i + h, \dots x_n) - f_1(x_1, x_2, \dots, x_n) / h \\ \lim_{h\to 0} f_2(x_1, \dots, x_i + h, \dots x_n) - f_1(x_1, x_2, \dots, x_n) / h \\ \vdots \\ \lim_{h\to 0} f_m(x_1, \dots, x_i + h, \dots x_n) - f_1(x_1, x_2, \dots, x_n) / h \end{bmatrix} = \begin{bmatrix} \partial f_1 / \partial x_i \\ \vdots \\ \partial f_m / \partial x_i \end{bmatrix}$$

$$\frac{d}{d\vec{x}} \vec{f}(\vec{x}) = \begin{bmatrix} \frac{\vec{f}(\vec{x})}{\partial x_1} & \frac{\vec{f}(\vec{x})}{\partial x_2} & \cdots & \frac{\vec{f}(\vec{x})}{\partial x_n} \end{bmatrix}$$

$$= \begin{bmatrix} \partial f_1/\partial x_1 & \partial f_1/\partial x_2 & \cdots & \partial f_1/\partial x_n \\ \partial f_2/\partial x_1 & \partial f_2/\partial x_2 & \cdots & \partial f_2/\partial x_n \\ \vdots & \vdots & & \vdots \\ \partial f_m/\partial x_1 & \partial f_m/\partial x_2 & \cdots & \partial f_m/\partial x_n \end{bmatrix}$$

$$\frac{\partial}{\partial B_{ij}} \overset{m\times n}{A}$$

$(m \times n) \times (p \times q)$

$\therefore J_{ij} = \frac{\partial f_i}{\partial x_j}$ and $J \in R^{m \times n}$

$$\vec{f}(\vec{x}) = A\vec{x} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1N} \\ \vdots & & & \\ A_{M1} & A_{M2} & \cdots & A_{MN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} A_{11}x_1 + A_{12}x_2 + \dots + A_{1N}x_N \\ A_{21}x_1 + A_{22}x_2 + \dots + A_{2N}x_N \\ \vdots \\ A_{M1}x_1 + \dots + A_{MN}x_N \end{bmatrix}$$

$$\frac{\partial}{\partial \vec{x}}(A\vec{x}) = \begin{bmatrix} \frac{\partial A\vec{x}}{\partial x_1} & \frac{\partial A\vec{x}}{\partial x_2} & \cdots & \frac{\partial A\vec{x}}{\partial x_N} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1N} \\ A_{21} & A_{22} & & A_{2N} \\ \vdots & \vdots & & \vdots \\ A_{M1} & A_{M2} & & A_{MN} \end{bmatrix} = A$$

$$\frac{\partial (A\vec{x})}{\partial x_i} = \begin{bmatrix} A_{1i} \\ A_{2i} \\ \vdots \\ A_{Mi} \end{bmatrix}$$

$f: \mathbb{R}^{M\times N} \to \mathbb{R}^p$

$$\begin{bmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_p(x) \end{bmatrix} \qquad \frac{\partial}{\partial x} f_K(x) \downarrow \frac{\partial}{\partial x_{ij}} f_K(x)$$

$$\frac{\partial f_K(x)}{\partial x} = \begin{bmatrix} \frac{\partial}{\partial x_{11}} f_K(x) & \frac{\partial}{\partial x_{12}} f_K(x) & \cdots & \frac{\partial f_K(x)}{\partial x_{1N}} \\ \frac{\partial f_K(x)}{\partial x_{M1}} & \cdots\cdots & & \frac{\partial f_K(x)}{\partial x_{MN}} \end{bmatrix}$$

$$\frac{\partial}{\partial \vec{x}}(\vec{x}^T\vec{a}) = \begin{bmatrix} \frac{\partial \vec{x}^T\vec{a}}{\partial x_1} & \frac{\partial \vec{x}^T\vec{a}}{\partial x_2} & \frac{\partial \vec{x}^T\vec{a}}{\partial x_3} & \cdots & \frac{\partial \vec{x}^T\vec{a}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}$$

$$= a^T.$$

$$\underset{1\times n}{\vec{x}^T} \underset{n\times 1}{\vec{a}} = [x_1 \; x_2 \cdots x_n] \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

$$\therefore \frac{\partial}{\partial \vec{x}}(\vec{x}^T\vec{a}) = \vec{a}^T$$

$$= a_1x_1 + a_2x_2 + \cdots a_nx_n$$

$$\frac{\partial}{\partial x_i}(\vec{x}^T\vec{a}) = a_i$$

$$\frac{\partial}{\partial \vec{x}}(\vec{a}^T x)$$

$x \to n\times 1$, $\vec{a}^T \to 1\times n \Rightarrow \vec{a} = \vec{n}\times 1$

$$\vec{a}^T\vec{x} = [a_1 \; a_2 \cdots a_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = a_1x_1 + \cdots + a_nx_n$$

$$\frac{\partial}{\partial x_i}(a^Tx) = a_i$$

$$\therefore \frac{\partial}{\partial \vec{x}}(\vec{a}^T\vec{x}) = \begin{bmatrix} \frac{\partial \vec{a}^T\vec{x}}{\partial x_1} & \frac{\partial \vec{a}^T\vec{x}}{\partial x_2} & \cdots & \frac{\partial \vec{a}^T\vec{x}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}$$

$$\therefore \frac{\partial}{\partial \vec{x}}(\vec{a}^T\vec{x}) = \vec{a}^T$$

$$\frac{\partial}{\partial X} a^T X b$$

let $X \rightarrow m \times n$, then $\vec{b} \rightarrow n \times 1$ & $\vec{a} \rightarrow m \times 1$

$\therefore \quad a^T X b \rightarrow 1 \times m, m \times n, n \times 1 \rightarrow (1)$

$$\vec{a}^T X \vec{b} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$$= \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \begin{bmatrix} x_{11} b_1 + x_{12} b_2 + x_{13} b_3 + \cdots x_{1n} b_n \\ x_{21} b_1 + x_{22} b_2 + x_{23} b_3 + \cdots x_{2n} b_n \\ \vdots \\ x_{m1} b_1 + x_{m2} b_2 + \cdots x_{mn} b_n \end{bmatrix}$$

$$= a_1 (x_{11} b_1 + x_{12} b_2 + x_{13} b_3 + \cdots x_{1n} b_n) + a_2 (x_{21} b_1 + x_{22} b_2 + x_{23} b_3 + \cdots x_{2n} b_n)$$

$$+ \cdots + a_n (x_{m1} b_1 + x_{m2} b_2 + \cdots x_{mn} b_n)$$

$$\frac{\partial}{\partial X_{ij}} (a^T X b) = a_i \cdot b_j$$

$$\frac{\partial}{\partial X} (a^T X b) = \begin{bmatrix} \frac{\partial}{\partial x_{11}} (a^T X b) & \frac{\partial}{\partial x_{12}} (a^T X b) & \cdots & \frac{\partial}{\partial x_{1n}} (a^T X b) \\ \vdots & & & \\ \frac{\partial}{\partial x_{m1}} (a^T X b) & \frac{\partial}{\partial x_{m2}} (a^T X b) & \cdots & \frac{\partial}{\partial x_{mn}} (a^T X b) \end{bmatrix}$$

$$= \begin{bmatrix} a_1 b_1 & a_1 b_2 & \cdots & a_1 b_n \\ a_2 b_1 & a_2 b_2 & \cdots & a_2 b_n \\ \vdots & & & \\ a_m b_1 & a_m b_2 & \cdots & a_m b_n \end{bmatrix}$$

$$= \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} \begin{bmatrix} b_1 & b_2 & \cdots & b_n \end{bmatrix} = a b^T$$

$x \to n \times 1$ $\quad B \to n \times n$

$x^T \to 1 \times n$

$$\frac{\partial}{\partial \vec{x}} (\vec{x}^T \overset{\text{mat}}{B} \vec{x}) = ? \qquad B\vec{x}$$

$$\vec{x}^T B \vec{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1n} \\ B_{21} & B_{22} & \cdots & B_{2n} \\ \vdots \\ B_{n1} & B_{n2} & \cdots & B_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$= \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} B_{11}x_1 + B_{12}x_2 + \cdots + B_{1n}x_n \\ B_{21}x_1 + B_{22}x_2 + \cdots B_{2n}x_n \\ \vdots \\ B_{n1}x_1 + B_{n2}x_2 + \cdots B_{nn}x_n \end{bmatrix}$$

$$= x_1(B_{11}x_1 + B_{12}x_2 + \cdots + B_{1n}x_n) + x_2(B_{21}x_1 + B_{22}x_2 + \cdots B_{2n}x_n) + x_n(B_{n1}x_1 + B_{n2}x_2 + \cdots B_{nn}x_n)$$

$$\therefore \frac{\partial}{\partial x_K}(x^T B x) = \frac{\partial}{\partial x_K}\left(\sum_{i=1}^{n} x_i (B_{i1}x_1 + B_{i2}x_2 + \cdots B_{in}x_n)\right)$$

$$= \frac{\partial}{\partial x_K}\left(\sum_{i=1}^{n} x_i \sum_{j=1}^{n} B_{ij} x_j\right)$$

$$= \frac{\partial}{\partial x_K}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} x_i B_{ij} x_j\right)$$

$$= \frac{\partial}{\partial x_K}\Big($$

$$\sum_{i=1}^{K-1} x_i (B_{i1}x_1 + \cdots B_{in}x_n) \to \sum_{i=1}^{K-1} x_i \cdot B_{iK}$$

$$+ x_K(B_{K1}x_1 + \cdots B_{Kn}x_n) \to 2B_{KK}x_K$$

$$+ \sum_{i=K+1}^{n} x_i (B_{i1}x_1 + \cdots + B_{in}x_n) \to \sum_{i=K+1}^{n} x_i \cdot B_{iK}$$

$\frac{\partial}{\partial \beta}(\beta^T X^T X \beta)$

$\frac{\partial}{\partial \beta}((X\beta)^T X\beta)$

$\left(\frac{\partial X\beta}{\partial \beta}\right)\left(\frac{\partial}{\partial X\beta}\right)((X\beta)^T X\beta)$

$X \to n \times 1$

$\beta \to n \times m$

LR
K-means
GMM
PCA

$\boxed{L \& M}$

$\boxed{6 \text{ hours}}$

$\boxed{12 \text{ hours}}$

$\boxed{4 \text{ hours}}$ $\boxed{2 \text{ hours}}$

$\boxed{6 \text{ hours}}$

$\frac{\partial}{\partial} f(\vec{v})$

$f : \mathbb{R}^n \to \mathbb{R}^m$

$\frac{\partial}{\partial \vec{v}} \vec{f}(\vec{x}) = ?$

$\Sigma_K \to 0$

$\theta = \begin{matrix} \mu_K \\ \Sigma_K \\ \pi_K \end{matrix} \quad k = 1, \dots, K$

$r_{1i}, r_{2i}, \dots r_{Ki} \in [0,1]$

$X^T X$

$\beta \quad m \times n$

$1 \times n \quad n \times 1$

$\beta^T X^T X \beta$

$(\beta_1 \; \beta_2 \dots \beta_n)\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}$

$\frac{\partial}{\partial \vec{x}}(f(x))$

KNN

r

$r_{Ki} = \frac{\pi_K N(x_i \mid \mu_K, \Sigma_K)}{\sum_{K=1}^{K} \pi_K N(x_i \mid \mu_K, \Sigma_K)}$

$\frac{\partial}{\partial \beta}(\beta^T \beta)$

$\sum_{i=1}^{n} \beta_i^2$

$\frac{\partial}{\partial \beta_i}\left(\sum_{i=1}^{n} \beta_i^2\right)$

$2\beta_i \qquad 2\beta$

$\frac{\partial}{\partial \beta}(\beta^T \beta) = 2\beta$

1) cal

2) $c_{new} = \frac{\sum x_i}{n}$

$\frac{\partial}{\partial \beta}(\beta^T X^T X \beta)$

$X^T X \frac{\partial}{\partial \beta}(\beta^T \beta)$

$\boxed{2 X^T X \beta}$

E

1) r_{Ki}

M

2) $\mu_{mean} = \frac{\sum \mu_i r_{Ki}}{\sum r_{Ki}} \to \mu_K$

$[0, 1, \dots, 0]$

$\downarrow$

$\Sigma = 0$

μ_K

Σ_i

$\Sigma^{(t+1)} = \frac{\sum_{K=1}^{K} r_{Ki}(X - \mu_K)(X - \mu_K)^T}{\sum_{K=1}^{K} r_{Ki}} = 0$

X

$\overbrace{0.2}\;\overbrace{0.5}\;\overbrace{0.3}$

$\pi_K = \pi_K$

$$\frac{\partial}{\partial x}(x^T A) = [x_1 \cdots x_m] \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

$\underset{1\times m}{} \; \underset{m\times n}{A}$

$$1\times n \quad = \left[a_{11}x_1 + a_{21}x_2 + \cdots a_{m1}x_m,\; a_{12}x_1 + a_{22}x_2 + \cdots + a_{m2}x_m, \cdots \right.$$

$$\left. a_{1n}x_1 + a_{2n}x_2 + \cdots a_{mn}x_m \right]$$

$$\frac{\partial}{\partial x_i}[x^T A] = [\; a_{i1} \quad a_{i2} \quad \cdots \quad a_{in} \;]$$

$$\frac{\partial}{\partial x}(x^T A) = \begin{bmatrix} \frac{\partial}{\partial x_1} x^T A \\ \frac{\partial}{\partial x_2} x^T A \\ \vdots \\ \frac{\partial}{\partial x_m} x^T A \end{bmatrix}$$

$$\frac{\partial}{\partial x}(x^T A) = A$$

$\nabla \to$ column vector

$$= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{m1} & a_{m2} & & a_{mn} \end{bmatrix} = A$$

$\nabla(s) = [\qquad]$

$- 2 x_{ik} x_{jk}$

$$\sum_{k=1}^{K} \frac{1}{n_k} \sum_{j=1}^{n_k} \sum_{i=1}^{n_k} (x_{ik} - x_{jk})^2$$

$$\sum_{k=1}^{K} \sum_{i=1}^{n_k} x_{ik}^2 - 2 \mu_k^2 n_k^2$$

$$2 n_k^2 \cdot \sum_{i=1}^{n_k} |x_{ik} - \mu_k|^2$$

PCA $$\Sigma = \frac{\sum_{i=1}^{n} (x_i - \mu)(x_i - \mu)^T}{n}$$

Let

$S = \frac{1}{n} \sum x x^T$ → data covariance matrix

$Var(z) = Var(B^T (x - \mu)) = Var(B^T x)$ assume $\mu \to 0$

$z_n = B^T x_n$

$\hat{x}_n = B B^T x_n = B z_n$

$$\frac{\partial}{\partial x} x^T A x = x^T (A + A^T)$$

$$Var(z_{1n}) = Var(b_1^T x_n)$$
$$= \frac{1}{N} \sum_{n=1}^{N} (z_{1n})^2$$
$$= \frac{1}{N} \sum_{n=1}^{N} b_1^T x_n x_n^T b_1$$
$$= \frac{1}{N} b_1^T S b_1$$

Maximize $Var(z_{1n})$ s.t. $b_1^T b_1 = 1$

$$\frac{\partial}{\partial b_1} (b_1^T S b_1) = \lambda \cdot \frac{\partial}{\partial b_1} (b_1^T b_1)$$
$$\cancel{2} b_1^T S = \lambda \cancel{2} b_1^T$$
$$\boxed{S b_1 = \lambda b_1}$$

$$g(z) = \frac{1}{1+e^{-z}} \quad \text{(sigmoid)}$$

$$h_\theta(x) = g(\theta^T x)$$

Back propogation

Neural Networks from scratch

Implementation design

for a modular code, we need to implement every layer seperately, in a different class

And for that, we need a model that fits every type of layer. Eg.

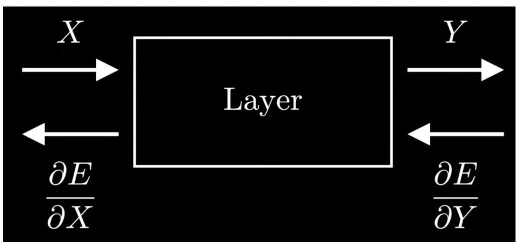
→ Also called dense
fully connected
convolutional
drop out
activation

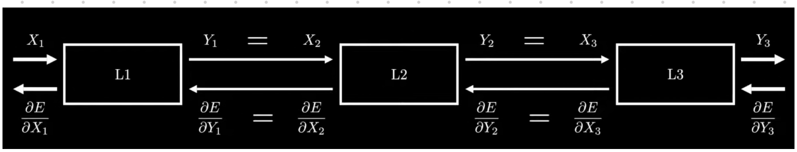
Every layer has an input and a output

This is called forward pass / forward pass.

The other step is backward propogation / backward pass when the layer updates its parameters

Given the derivative/gradient of the error wrt to the input, the layer should give the derivative/gradient wrt. to its input



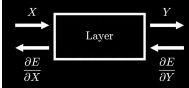


i) why would we need $\frac{\partial E}{\partial Y}$? Because using that, we can update the parameters by calculating the gradient wrt. to weights of the layer as follows: $\frac{\partial E}{\partial w} = \frac{\partial Y}{\partial w} \cdot \frac{\partial E}{\partial Y}$, and since $\frac{\partial Y}{\partial w}$ is Local, it can be calculated.

ii) why would we need $\frac{\partial E}{\partial X}$? Because the input to the current layer would be the the output of the previous layer, and therefore, giving that as an input to the previous layer would allow us to calculate gradients wrt to its parameters & update them.

This will happen until every layer is capable of updating its parameters.

Base Layer




```python
class Layer:
    def __init__(self):
        self.input = None
        self.output = None
#
    def forward(self, input):
        # TODO: return output
        pass
#
    def backward(self, output_gradient, learning_rate):
        # TODO: update parameters and return input gradient
        pass
```

here for convinience so other layers can inherit and do not have to declare them

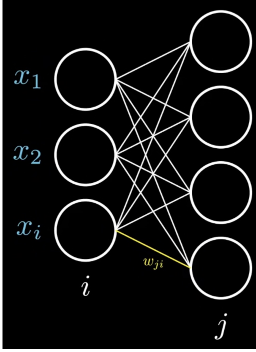
$\partial E/\partial Y$

some constant

calculating $\partial E/\partial w$, $\partial E/\partial b$

$\partial E/\partial x$

Dense Layer



$$y_1 = x_1 w_{11} + x_2 w_{12} + \cdots + x_i w_{1i} + b_1$$

$$y_2 = x_1 w_{21} + x_2 w_{22} + \cdots + x_i w_{2i} + b_2$$

$$y_3 = x_1 w_{31} + x_2 w_{32} + \cdots + x_i w_{3i} + b_3$$

$$y_j = x_1 w_{j1} + x_2 w_{j2} + \cdots + x_i w_{ji} + b_j$$

This is forward prop for dense layer

$$Y = W \cdot X + B$$

$$\begin{cases} y_1 = x_1 w_{11} + x_2 w_{12} + \cdots + x_i w_{1i} + b_1 \\ y_2 = x_1 w_{21} + x_2 w_{22} + \cdots + x_i w_{2i} + b_2 \\ y_3 = x_1 w_{31} + x_2 w_{32} + \cdots + x_i w_{3i} + b_3 \\ y_j = x_1 w_{j1} + x_2 w_{j2} + \cdots + x_i w_{ji} + b_j \end{cases} \Leftrightarrow \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_j \end{bmatrix}}_{j \times 1} = \underbrace{\begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1i} \\ w_{21} & w_{22} & \cdots & w_{2i} \\ \vdots & \vdots & \ddots & \vdots \\ w_{j1} & w_{j2} & \cdots & w_{ji} \end{bmatrix}}_{j \times i} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \end{bmatrix}}_{i \times 1} + \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_j \end{bmatrix}}_{j \times 1}$$

$$Y = W \cdot X + B$$

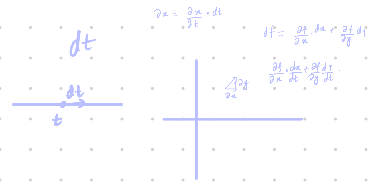
```python
from layer import Layer
import numpy as np
#
class Dense(Layer):
    def __init__(self, input_size, output_size):
        self.weights = np.random.randn(output_size, input_size)
        self.bias = np.random.randn(output_size, 1)
#
    def forward(self, input):
        self.input = input
        return np.dot(self.weights, self.input) + self.bias
#
    def backward(self, output_gradient, learning_rate):
        # TODO: update parameters and return input gradient
        pass
```

$Y = WX + B$

$x = (f, n)$

$(j \times i) \times (i, n)$

$j \times n$



$$\frac{\partial E}{\partial Y} \to \frac{\partial E}{\partial W}, \frac{\partial E}{\partial B}$$

needed for updation of parameters of the current layer

$$\frac{\partial E}{\partial Y} \to \frac{\partial E}{\partial X}$$

Needed to pass it to previous layer for it's updation

Given this

$$\frac{\partial E}{\partial Y} = \underbrace{\begin{bmatrix} \frac{\partial E}{\partial y_1} \\ \frac{\partial E}{\partial y_2} \\ \vdots \\ \frac{\partial E}{\partial y_j} \end{bmatrix}}_{j \times 1}$$

$\rightarrow$

we need this

$$\frac{\partial E}{\partial W} = \underbrace{\begin{bmatrix} \frac{\partial E}{\partial w_{11}} & \frac{\partial E}{\partial w_{12}} & \cdots & \frac{\partial E}{\partial w_{1i}} \\ \frac{\partial E}{\partial w_{21}} & \frac{\partial E}{\partial w_{22}} & \cdots & \frac{\partial E}{\partial w_{2i}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial E}{\partial w_{j1}} & \frac{\partial E}{\partial w_{j2}} & \cdots & \frac{\partial E}{\partial w_{ji}} \end{bmatrix}}_{j \times i}$$

$$\frac{\partial E}{\partial w_{12}} = \frac{\partial E}{\partial y_1} \cdot \frac{\partial y_1}{\partial w_{12}} + \frac{\partial E}{\partial y_2} \frac{\partial y_2}{\partial w_{12}} + \cdots + \frac{\partial E}{\partial y_j} \cdot \frac{\partial y_j}{\partial w_{12}}$$

Multivariable chain rule

$$\frac{\partial E}{\partial w_{12}} = x_2 \cdot \frac{\partial E}{\partial y_1}$$

$$\therefore \frac{\partial E}{\partial w_{ij}} = x_j \cdot \frac{\partial E}{\partial y_i}$$

$$\boxed{\frac{\partial E}{\partial W} = \frac{\partial E}{\partial Y} \cdot X^T}$$

$$\frac{\partial E}{\partial W} = \underbrace{\begin{bmatrix} x_1 \frac{\partial E}{\partial y_1} & x_2 \frac{\partial E}{\partial y_1} & \cdots & x_i \frac{\partial E}{\partial y_1} \\ x_1 \frac{\partial E}{\partial y_2} & x_2 \frac{\partial E}{\partial y_2} & \cdots & x_i \frac{\partial E}{\partial y_2} \\ \\ x_1 \frac{\partial E}{\partial y_j} & x_2 \frac{\partial E}{\partial y_j} & \cdots & x_i \frac{\partial E}{\partial y_j} \end{bmatrix}}_{j \times i}$$

$$= \begin{bmatrix} \partial E / \partial y_1 \\ \vdots \\ \partial E / \partial y_j \end{bmatrix} \begin{bmatrix} x_1 & x_2 & \cdots & x_i \end{bmatrix} = \frac{\partial E}{\partial Y} X^T$$

One final output $f(x(t), y(t))$

Two intermediate outputs $x(t)$ $y(t)$

One input t

There is still a chain rule that lets you compute the derivative of this new single-variable function $f(x(t), y(t))$, and it involves the partial derivatives of f :

$$\underbrace{\frac{d}{dt}}_{\text{This is an ordinary derivative, not a partial derivative } \frac{\partial}{\partial t} \text{ because the total composition has one input and one output.}} f(x(t), y(t)) = \underbrace{\overbrace{\frac{\partial f}{\partial x}}^{\text{How } f \text{ changes due to a tiny change in } x} \overbrace{\frac{dx}{dt}}^{\text{How } x \text{ changes due to a tiny change in } t}}_{\text{Total change in } f \text{ due to the influence } t \text{ has on } x} + \underbrace{\frac{\partial f}{\partial y} \frac{dy}{dt}}_{\text{Total change in } f \text{ due to the influence } t \text{ has on } y}$$

$$\frac{\partial E}{\partial b_2} = \overbrace{\frac{\partial y_1}{\partial b_2}}^{0}\frac{\partial E}{\partial y_1} + \overbrace{\frac{\partial y_2}{\partial b_2}}^{1}\cdot\frac{\partial E}{\partial y_2} + \dots + \overbrace{\frac{\partial y_j}{\partial b_2}}^{0}\cdot\frac{\partial E}{\partial y_j}$$

$$= \frac{\partial E}{\partial y_2}$$

$$\frac{\partial E}{\partial b_2} = \frac{\partial E}{\partial y_2} \Rightarrow \boxed{\frac{\partial E}{\partial B} = \frac{\partial E}{\partial Y}}$$

$$\frac{\partial E}{\partial x_2} = \frac{\partial y_1}{\partial x_2}\cdot\frac{\partial E}{\partial y_1} + \frac{\partial y_2}{\partial x_2}\cdot\frac{\partial E}{\partial y_2} + \frac{\partial y_3}{\partial x_2}\cdot\frac{\partial E}{\partial y_3} + \dots + \frac{\partial y_j}{\partial x_2}\frac{\partial E}{\partial y_j}$$

$$\frac{\partial E}{\partial x_2} = w_{12}\frac{\partial E}{\partial y_1} + w_{22}\frac{\partial E}{\partial y_2} + w_{32}\frac{\partial E}{\partial y_3} + \dots + w_{j2}\frac{\partial E}{\partial y_j}$$

$$\Rightarrow \frac{\partial E}{\partial x_i} = w_{1i}\frac{\partial E}{\partial y_1} + w_{2i}\frac{\partial E}{\partial y_2} + \dots + w_{ji}\frac{\partial E}{\partial y_j}.$$

$$\frac{\partial E}{\partial X} = \begin{bmatrix} \vdots \\ \sum_{j=1}^{j} w_{ji}\frac{\partial E}{\partial y_j} \\ \vdots \end{bmatrix}_{i\times 1}$$

$$= \quad i\times j \quad j\times 1$$

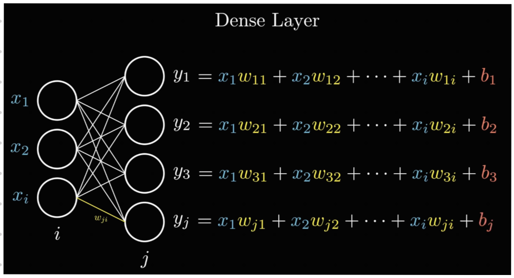
$$= \begin{bmatrix} w_{11} & w_{21} & w_{31} & \dots & w_{i1} \\ \vdots & & & & \\ w_{1j} & w_{2j} & \dots & & w_{ij} \end{bmatrix}$$

$$E = H(y, \hat{y})$$

$$= -\sum_{i=1}^{c} y_i^* \log y_i$$

$$\frac{\partial E}{\partial y_1} = \frac{-y_1^*}{y_1}$$

$$\frac{\partial E}{\partial Y} = -\begin{bmatrix} y_1^* \\ y_2^* \\ \vdots \\ y_c^* \end{bmatrix} \cdot$$



$$X \to Y$$

$$\frac{\partial E}{\partial X} \leftarrow \frac{\partial E}{\partial Y}$$

$\hookrightarrow (n, f)$

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1f} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2f} \\ \vdots \\ x_{n1} & x_{n2} & x_{n3} & \cdots & x_{nf} \end{bmatrix}_{n \times f} = X \to X^T$$

$$y_{11} =$$

$$W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1i} \\ w_{21} & w_{22} & \cdots & w_{2i} \\ \vdots \\ w_{j1} & w_{j2} & \cdots & w_{ji} \end{bmatrix}_{j \times f}$$

$$\vec{y}_1 = \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1j} \end{bmatrix}$$

$$[y_1 \quad y_2 \quad \cdots \quad y_n]$$

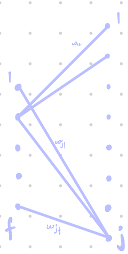
$$\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ \vdots \\ y_{j1} & y_{j2} & \cdots & y_{jn} \end{bmatrix} = Y$$

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1n} \\ x_{21} \\ x_{31} \\ \vdots \\ x_{f1} & & & & x_{fn} \end{bmatrix}_{f \times n}$$

$$j \times n$$

$$WX + B$$

$w_{12} \quad w_{22} \quad \cdots \quad w_{j2}$



w_{jf}

$$wx = [y_1 \quad y_2 \quad \cdots \quad y_n]$$

$$y_1 = \Big[\, x_{11} w_{11} + x_{21} w_{12} + \cdots \; x_{f1} w_{1f}$$

$$W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1f} \\ w_{21} & w_{22} & \cdots & w_{2f} \\ \vdots & & & \\ w_{j1} & w_{j2} & \cdots & w_{jf} \end{bmatrix}_{j\times f} = \qquad X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1f} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2f} \\ \vdots & & & & \\ x_{n1} & x_{n2} & x_{n3} & \cdots & x_{nf} \end{bmatrix}_{n\times f}$$

$$Y = WX^T = (j, n)$$

$$XW^T = (n, j) \ (\checkmark)$$

$$XW^T = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}_{n\times j} + \begin{bmatrix} b_1 & b_2 & \cdots & b_j \\ b_1 & b_2 & & b_j \\ \vdots & \to n \text{ times} & & \\ b_1 & b_2 & \cdots & b_j \end{bmatrix}$$

$$\boxed{w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_j \end{bmatrix}}$$

$$\boxed{Y = XW^T + B}$$ → forward propagation equation

Given $\Rightarrow \dfrac{\partial E}{\partial Y}$, find $\dfrac{\partial E}{\partial X}$

$$\frac{\partial E}{\partial X} = \begin{bmatrix} \partial E/\partial x_{11} & \partial E/\partial x_{12} & \partial E/\partial x_{13} & \cdots & \partial E/\partial x_{1f} \\ \partial E/\partial x_{21} & \partial E/\partial x_{22} & & \cdots & \partial E/\partial x_{2f} \\ \vdots & & & & \\ \partial E/\partial x_{n1} & \partial E/\partial x_{n2} & & \cdots & \partial E/\partial x_{nf} \end{bmatrix}^T = \begin{bmatrix} \partial E/\partial x_1 \\ \partial E/\partial x_2 \\ \vdots \\ \partial E/\partial x_n \end{bmatrix}^T_{f\times n}$$

$$y_1 = [y_{11} \ y_{12} \ y_{13} \cdots y_{1j}] \qquad Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} \Rightarrow \frac{\partial E}{\partial Y} = \left[\frac{\partial E}{\partial y_1} \ \frac{\partial E}{\partial y_2} \cdots \frac{\partial E}{\partial y_n} \right]_{j\times n}$$

$$y_{11} = x_{11} w_{11} + x_{12} w_{12} + \cdots + x_{1f} w_{1f} + b_1$$

$$y_{12} = x_{11} w_{21} + x_{12} w_{22} + \cdots + x_{1f} w_{2f} + b_2$$

$$y_{13} = x_{11} w_{31} + x_{12} w_{32} + \cdots + x_{1f} w_{3f} + b_3$$

$$\vdots$$

$$y_{1j} = x_{11} w_{j1} + x_{12} w_{j2} + \cdots + x_{1f} w_{jf} + b_j$$

$$\frac{\partial E}{\partial x_{12}} = \frac{\partial y_{11}}{\partial x_{12}} \cdot \frac{\partial E}{\partial y_{11}} + \frac{\partial y_{12}}{\partial x_{12}} \cdot \frac{\partial E}{\partial y_{12}} + \cdots + \frac{\partial y_{1j}}{\partial x_{12}} \cdot \frac{\partial E}{\partial y_{1j}}$$

$$= w_{12} \frac{\partial E}{\partial y_{11}} + w_{22} \frac{\partial E}{\partial y_{12}} + \cdots + w_{j2} \frac{\partial E}{\partial y_{1j}}$$

$$\frac{\partial E}{\partial x_{\ell k}} = \sum_{p=1}^{j} w_{pk} \frac{\partial E}{\partial y_{\ell p}} =$$

$$W = \begin{bmatrix} w_1 & w_2 & \cdots & w_f \end{bmatrix}_{j \times f}$$

$$\frac{\partial E}{\partial W} = \begin{bmatrix} \partial E / \partial w_1 \\ \partial E / \partial w_2 \\ \vdots \\ \partial E / \partial w_f \end{bmatrix}$$

$$\frac{\partial E}{\partial x_\ell} = (f, 1), \quad w_k = (1, j) \quad \frac{\partial E}{\partial y_\ell} = (j, 1)$$

$$\frac{\partial E}{\partial x_{\ell k}} = w_k \frac{\partial E}{\partial y_\ell}$$

$$\frac{\partial E}{\partial X} = \begin{bmatrix} w_1 \frac{\partial E}{\partial y_1} & w_2 \frac{\partial E}{\partial y_1} & w_3 \frac{\partial E}{\partial y_1} & \cdots & w_f \frac{\partial E}{\partial y_1} \\ w_1 \frac{\partial E}{\partial y_2} & w_2 \frac{\partial E}{\partial y_2} & w_3 \frac{\partial E}{\partial y_2} & \cdots & w_f \frac{\partial E}{\partial y_2} \\ \vdots & & & & \\ w_1 \frac{\partial E}{\partial y_n} & w_2 \frac{\partial E}{\partial y_n} & w_3 \frac{\partial E}{\partial y_n} & \cdots & w_f \frac{\partial E}{\partial y_n} \end{bmatrix}^T_{f \times n}$$

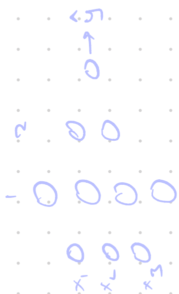
$$= \left(\frac{\partial E}{\partial Y} \cdot W \right)^T$$

$$= \underset{f \times j}{W^T} \cdot \frac{\partial E}{\partial Y}_{j \times n}$$

$$\frac{\partial E}{\partial Y} = \begin{bmatrix} \frac{\partial E}{\partial y_1} & \frac{\partial E}{\partial y_2} & \cdots & \frac{\partial E}{\partial y_n} \end{bmatrix}_{j \times n}$$

$$\boxed{\frac{\partial E}{\partial X} = W^T \cdot \frac{\partial E}{\partial Y}}$$

$$\frac{\partial E}{\partial w_{12}} = \frac{1}{n} \sum_{i=1}^{n} \boxed{\sum_{k=1}^{j} \frac{\partial y_{ik}}{\partial w_{12}} \cdot \frac{\partial E}{\partial y_{ik}}}$$



$$\frac{\partial E}{\partial W} = n \times j \times f$$

$j \times f$

for x_1

$$\frac{\partial E}{\partial w_{12}} = \frac{\partial y_{11}}{\partial w_{12}} \cdot \frac{\partial E}{\partial y_{11}} + \frac{\partial y_{12}}{\partial w_{12}} \cdot \frac{\partial E}{\partial y_{12}} + \dots + \frac{\partial y_{1j}}{\partial w_{12}} \cdot \frac{\partial E}{\partial y_{1j}}$$

$$= x_{12} \frac{\partial E}{\partial y_{11}}$$

$$y_{11} = x_{11} w_{11} + x_{12} w_{12} + \dots + x_{1f} w_{1f}$$
$$y_{12} = x_{11} w_{21} + x_{12} w_{22} + \dots + x_{1f} w_{2f}$$
$$y_{13} = x_{11} w_{31} + x_{12} w_{32} + \dots + x_{1f} w_{3f}$$
$$\vdots$$
$$y_{1j} = x_{11} w_{j1} + x_{12} w_{j2} + \dots + x_{1f} w_{jf}$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}_{n \times j} \quad X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1f} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2f} \\ \vdots \\ x_{n1} & x_{n2} & x_{n3} & \dots & x_{nf} \end{bmatrix}_{n \times f} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

for x_1,

$$\frac{\partial E}{\partial w_{lk}} = x_{1k} \frac{\partial E}{\partial y_{1l}} \quad \Bigg| \quad \frac{\partial E}{\partial w} = \left[\frac{\partial E}{\partial w_1} \ \frac{\partial E}{\partial w_2} \ \dots \ \frac{\partial E}{\partial w_j} \right]$$

$\therefore$ for x_{1},

$$\frac{\partial E}{\partial w} = \begin{bmatrix} \partial E/\partial w_{11} & \partial E/\partial w_{21} & & \partial E/\partial w_{j1} \\ \partial E/\partial w_{12} & \partial E/\partial w_{22} & \dots & \partial E/\partial w_{j2} \\ \vdots & \vdots & & \vdots \\ \partial E/\partial w_{1f} & \partial E/\partial w_{2f} & & \partial E/\partial w_{jf} \end{bmatrix}_{f \times j} = \begin{bmatrix} x_{11}\, \partial E/\partial y_{11} & x_{11}\, \partial E/\partial y_{12} & \dots & x_{11}\, \partial E/\partial y_{1j} \\ \vdots \\ x_{1f}\, \partial E/\partial y_{11} & x_{1f}\, \partial E/\partial y_{12} & \dots & x_{1f}\, \partial E/\partial y_{1j} \end{bmatrix}$$

$$= \begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1f} \end{bmatrix} \left[\frac{\partial E}{\partial y_{11}} \ \frac{\partial E}{\partial y_{12}} \ \dots \ \frac{\partial E}{\partial y_{1j}} \right]$$

$$= \underset{f \times 1}{x_1^T} \left(\frac{\partial E}{\partial y_{1_{(1,j)}}} \right)^T$$

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\updownarrow$$

$$\frac{\partial E}{\partial w} = x_1^T \frac{\partial E}{\partial y_1}$$ for x_1. let this be $\frac{\partial E}{\partial w^{(1)}}$ $\quad A \odot B$

Kind of Hadamard

$$\frac{\partial E}{\partial W} = \begin{bmatrix} \partial E/\partial w^{(1)} \\ \partial E/\partial w^{(2)} \\ \vdots \\ \partial E/\partial w^{(m)} \end{bmatrix}_{nj \times f} = \begin{bmatrix} \left(\frac{\partial E}{\partial y_1}\right)^T \cdot x_1 \\ \left(\frac{\partial E}{\partial y_2}\right)^T \cdot x_2 \\ \vdots \\ \left(\frac{\partial E}{\partial y_n}\right)^T \cdot x_n \end{bmatrix} = \begin{bmatrix} \left(\frac{\partial E}{\partial y_1}\right)^T \\ \left(\frac{\partial E}{\partial y_2}\right)^T \\ \vdots \\ \left(\frac{\partial E}{\partial y_n}\right)^T \end{bmatrix}_{nj \times 1} \odot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}_{n \times f} \quad [\quad]$$

∂E

$$= \frac{\partial E}{\partial Y} \odot X$$

$$\underset{1 \times 1}{\frac{\partial E}{\partial x_2}} = \left(\frac{\partial Y}{\partial x_2} \right)_{j \times 1} \cdot \frac{\partial E}{\partial Y}_{j \times 1}$$

$$= \left(\frac{\partial Y}{\partial x_2} \right)^T \cdot \frac{\partial E}{\partial Y}$$

$$= \left(\frac{\partial y_1}{\partial x_2} \ \frac{\partial y_2}{\partial x_2} \ \frac{\partial y_3}{\partial x_2} \ \dots \ \frac{\partial y_j}{\partial x_2} \right) \begin{bmatrix} \partial E/\partial y_1 \\ \partial E/\partial y_2 \\ \vdots \\ \partial E/\partial y_j \end{bmatrix}$$

Definitions:

$\vec{a} = [a_1 \ a_2 \cdots a_n]$

$$\frac{\partial c}{\partial a} = \begin{bmatrix} \partial c/\partial a_1 \\ \partial c/\partial a_2 \\ \vdots \\ \partial c/\partial a_n \end{bmatrix}$$

$X = n \times f, \ \frac{\partial E}{\partial Y} = j \times n$

$\frac{\partial E}{\partial w} = f \times j$

$$\sum \frac{\partial E}{\partial w} = X^T \frac{\partial E}{\partial Y}^T$$

$$= [x_1^T \ x_2^T \cdots x_n^T] \begin{bmatrix} \left(\frac{\partial E}{\partial y_1}\right)^T \\ \vdots \\ \left(\frac{\partial E}{\partial y_n}\right)^n \end{bmatrix}$$

$$=$$

$$\boxed{\left\langle \frac{\partial E}{\partial w} \right\rangle = \frac{1}{n} X^T \left(\frac{\partial E}{\partial Y}\right)^T}$$

$\boxed{\text{updated}}$

$\boxed{\text{incorrect}}$

$$\boxed{\left\langle \frac{\partial E}{\partial w} \right\rangle = \frac{1}{n} \left(\frac{\partial E}{\partial Y}\right)^T \cdot X \qquad \left\langle \frac{\partial E}{\partial w} \right\rangle = (j,n)\cdot(n,f)}$$

$$\begin{bmatrix} b_1 & b_2 & \cdots & b_j \\ b_1 & b_2 & & b_j \\ \vdots & n \text{ times} & & \\ b_1 & b_2 & \cdots & b_j \end{bmatrix}_{n \times j}$$

$b = [b_1 \ b_2 \cdots b_j]$

$$\frac{\partial E}{\partial b} = \begin{bmatrix} \partial E/\partial b_1 \\ \partial E/\partial b_2 \\ \vdots \\ \partial E/\partial b_j \end{bmatrix} = \begin{bmatrix} \partial E/\partial y_{11} \\ \partial E/\partial y_{12} \\ \vdots \\ \partial E/\partial y_{1j} \end{bmatrix}$$

$$= \frac{\partial E}{\partial y_1}$$

$\boxed{Y = XW^T + B}$ → forward propagation equation

Given $\frac{\partial E}{\partial Y}$, find $\frac{\partial E}{\partial B}$

$$\frac{\partial E}{\partial b_1} = \frac{\partial y_1}{\partial b_1}\frac{\partial E}{\partial y_1}$$

$$= \left[\frac{\partial y_{11}}{\partial b_1} \ \frac{\partial y_{12}}{\partial b_1} \cdots \frac{\partial y_{1j}}{\partial b_1} \right] \begin{bmatrix} \partial E/\partial y_{11} \\ \partial E/\partial y_{12} \\ \vdots \\ \partial E/\partial y_{1j} \end{bmatrix} = \frac{\partial E}{\partial y_{11}}$$

$$\frac{\partial E}{\partial B} = \begin{bmatrix} \partial E/\partial y_{11} & \partial E/\partial y_{12} & \cdots & \partial E/\partial y_{1j} \\ \vdots \rightarrow \text{n-times} & & & \\ \partial E/\partial y_{11} & \partial E/\partial y_{12} & \cdots & \partial E/\partial y_{1j} \end{bmatrix} \qquad B = \begin{bmatrix} - & b & - \\ - & b & - \\ & \vdots & \\ - & b & - \end{bmatrix}$$

$$= \begin{bmatrix} - & \partial E/\partial y_1 & - \\ - & \partial E/\partial y_1 & - \\ & \vdots & \\ - & \partial E/\partial y_1 & - \end{bmatrix}_{n \times j} \quad \text{for } y_1 \qquad B = \underbrace{\begin{bmatrix} b \\ b \\ \vdots \\ b \end{bmatrix}}_{\rightarrow n \text{ times}}$$

$$\frac{\partial E}{\partial B} = \left[\frac{\partial E}{\partial b} \; \frac{\partial E}{\partial b} \cdots \frac{\partial E}{\partial b}\right]$$

$$= \left[\frac{\partial E}{\partial y_1} \; \frac{\partial E}{\partial y_1} \cdots \frac{\partial E}{\partial y_1}\right]_{j \times n} \quad \text{for } y_1$$

$$\frac{\partial E}{\partial b} = \frac{\partial E}{\partial y_1} \cdot \frac{\partial y_1}{\partial b} \rightarrow \begin{bmatrix} \partial y_1/\partial b_1 \\ \partial y_1/\partial b_2 \\ \vdots \\ \partial y_1/\partial b_j \end{bmatrix}_{j \times j}$$

$$= \frac{\partial E}{\partial y_1} + \frac{\partial E}{\partial y_2} + \cdots \partial E/\partial y_n \rightarrow 1 \times j$$

Incorrect

$$\left\langle \frac{\partial E}{\partial B} \right\rangle = \frac{1}{n} \begin{bmatrix} \frac{\partial E}{\partial y_1} + \frac{\partial E}{\partial y_2} + \cdots \partial E/\partial y_n \\ \vdots \rightarrow \text{n-times} \\ \frac{\partial E}{\partial y_1} + \frac{\partial E}{\partial y_2} + \cdots \partial E/\partial y_n \end{bmatrix} = \text{mean rows of } \left|\frac{\partial E}{\partial Y}\right| \text{ tiled n-times}$$

Correct

$$\left\langle \frac{\partial E}{\partial B} \right\rangle = \frac{1}{n} \left[\sum_{i=1}^{n} \frac{\partial E}{\partial y_i} \;\; \sum_{i=1}^{n} \frac{\partial E}{\partial y_i} \;\; \overset{\boxed{\text{n-times}}}{\cdots} \;\; \sum_{i=1}^{n} \frac{\partial E}{\partial y_i} \right] = \text{mean of } \frac{\partial E}{\partial Y} \text{ across coloumn tiled n times along coloumn}$$

$$\frac{\partial E}{\partial Y} = \left[\frac{\partial E}{\partial y_1} \; \frac{\partial E}{\partial y_2} \cdots \frac{\partial E}{\partial y_n}\right]$$

$n \times f \longrightarrow n \times f$

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1f} \\ x_{2} & x_{21} & \cdots & x_{2f} \\ \vdots & & & \\ x_{n1} & x_{n2} & \cdots & x_{nf} \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \longrightarrow \begin{bmatrix} sf(x_1) \\ sf(x_2) \\ \vdots \\ sf(x_n) \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\frac{\partial E}{\partial Y} \rightarrow \left[\frac{\partial E}{\partial y_1} \quad \frac{\partial E}{\partial y_2} \quad \cdots \quad \frac{\partial E}{\partial y_n} \right]_{f \times n} \quad \text{Given}$$

find $$\frac{\partial E}{\partial X} = \left[\frac{\partial E}{\partial x_1} \quad \frac{\partial E}{\partial x_2} \quad \cdots \quad \frac{\partial E}{\partial x_n} \right]_{f \times n}$$

$$\frac{\partial E}{\partial x_{1k}} = \sum_{i=1}^{f} \frac{\partial E}{\partial y_{1i}} \cdot \frac{\partial y_{1i}}{\partial x_{1k}}$$

$$y_{1i} = \frac{e^{x_{1i}}}{\sum_{j=1}^{f} e^{x_{1j}}}$$

If $k = i$,

$$\frac{\partial y_{1i}}{\partial x_{1k}} = y_{1i} \cdot (1 - y_{1i})$$

If $k \neq i$

$$\frac{\partial y_{1i}}{\partial x_{1k}} = -y_{1i}\, y_{1k}$$

Valid for any sample j, not just 1

$$\frac{\partial y_{ji}}{\partial x_{jk}} = y_{ji}(1 - y_{ji}), \quad i = k$$

$$\frac{\partial y_{ji}}{\partial x_{jk}} = -y_{ji}\, y_{jk}$$

$$\frac{\partial E}{\partial X} = \begin{bmatrix} \partial E/\partial x_{11} & \partial E/\partial x_{12} & \cdots & \partial E/\partial x_{1f} \\ \vdots & & & \\ \partial E/\partial x_{n1} & \partial E/\partial x_{n2} & \cdots & \partial E/\partial x_{nf} \end{bmatrix}^T = \begin{bmatrix} \sum_{i=1}^{f} \frac{\partial E}{\partial y_{1i}} \cdot \frac{\partial y_{1i}}{\partial x_{11}} & \sum_{i=1}^{f} \frac{\partial E}{\partial y_{1i}} \cdot \frac{\partial y_{1i}}{\partial x_{12}} & \cdots & \sum_{i=1}^{f} \frac{\partial E}{\partial y_{1i}} \frac{\partial y_{1i}}{\partial x_{1f}} \\ \vdots & & & \\ \sum_{i=1}^{f} \frac{\partial E}{\partial y_{ni}} \cdot \frac{\partial y_{ni}}{\partial x_{n1}} & \sum_{i=1}^{f} \frac{\partial E}{\partial y_{ni}} \cdot \frac{\partial y_{ni}}{\partial x_{n2}} & \cdots & \sum_{i=1}^{f} \frac{\partial E}{\partial y_{ni}} \frac{\partial y_{ni}}{\partial x_{nf}} \end{bmatrix}^T$$

$$\frac{\partial E}{\partial X} = \begin{bmatrix} \partial E/\partial x_{11} & \partial E/\partial x_{12} & \cdots & \partial E/\partial x_{1f} \\ \vdots & & & \\ \partial E/\partial x_{n1} & \partial E/\partial x_{n2} & \cdots & \partial E/\partial x_{nf} \end{bmatrix}^T = \begin{bmatrix} \sum_{i=1}^{f} \frac{\partial E}{\partial y_{1i}} \cdot \frac{\partial y_{1i}}{\partial x_{11}} & \sum_{i=1}^{f} \frac{\partial E}{\partial y_{1i}} \cdot \frac{\partial y_{1i}}{\partial x_{12}} & \cdots & \sum_{i=1}^{f} \frac{\partial E}{\partial y_{1i}} \frac{\partial y_{1i}}{\partial x_{1f}} \\ \vdots & & & \\ \sum_{i=1}^{f} \frac{\partial E}{\partial y_{ni}} \cdot \frac{\partial y_{ni}}{\partial x_{n1}} & \sum_{i=1}^{f} \frac{\partial E}{\partial y_{ni}} \cdot \frac{\partial y_{ni}}{\partial x_{n2}} & \cdots & \sum_{i=1}^{f} \frac{\partial E}{\partial y_{ni}} \frac{\partial y_{ni}}{\partial x_{nf}} \end{bmatrix}$$

$$\frac{\partial y_1}{\partial x_{12}} = \frac{\partial}{\partial x_{12}} [\, y_{11} \;\; y_{12} \; \cdots \; y_{1f} \,]$$
$$= \begin{bmatrix} \frac{\partial y_{11}}{\partial x_{12}} & \frac{\partial_{12}}{\partial x_{12}} & \cdots & \frac{y_{1f}}{\partial x_{12}} \end{bmatrix}$$

$$\frac{\partial y_1}{\partial x_1} = \begin{bmatrix} \partial y_1/\partial x_{11} \\ \partial y_2/\partial x_{12} \\ \vdots \\ \partial y_n/\partial x_{1f} \end{bmatrix} = \begin{bmatrix} \frac{\partial y_{11}}{\partial x_{11}} & \frac{\partial y_{12}}{\partial x_{11}} & \cdots & \frac{\partial y_{1f}}{\partial x_{11}} \\ \vdots & & & \\ \frac{\partial y_{11}}{\partial x_{1f}} & \frac{\partial y_{12}}{\partial x_{1f}} & \cdots & \frac{\partial y_{1f}}{\partial x_{1f}} \end{bmatrix} = \begin{bmatrix} y_{11}(1-y_{11}) & -y_{12}y_{11} & \cdots & -y_{1f}y_{11} \\ -y_{11}y_{12} & y_{12}(1-y_{12}) & \cdots & -y_{1f}y_{12} \\ \vdots & \vdots & & \vdots \\ -y_{11}y_{1f} & -y_{12}y_{1f} & \cdots & y_{1f}(1-y_{1f}) \end{bmatrix}$$

$$\frac{\partial E}{\partial x_1} = \underset{f\times f}{\frac{\partial y_1}{\partial x_1}} \; \underset{f\times 1 \,\to\, f\times 1}{\frac{\partial E}{\partial y_1}} = \begin{bmatrix} \frac{\partial y_{11}}{\partial x_{11}} & \frac{\partial y_{12}}{\partial x_{11}} & \cdots & \frac{\partial y_{1f}}{\partial x_{11}} \\ \vdots & & & \\ \frac{\partial y_{11}}{\partial x_{1f}} & \frac{\partial y_{12}}{\partial x_{1f}} & \cdots & \frac{\partial y_{1f}}{\partial x_{1f}} \end{bmatrix} \begin{bmatrix} \partial E/\partial y_{11} \\ \partial E/\partial y_{12} \\ \vdots \\ \partial E/\partial y_{1f} \end{bmatrix}$$

$$\frac{\partial E}{\partial X} = \begin{bmatrix} \frac{\partial E}{\partial x_1} & \frac{\partial E}{\partial x_2} & \frac{\partial E}{\partial x_3} & \cdots & \frac{\partial E}{\partial x_n} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial y_1}{\partial x_1} \cdot \frac{\partial E}{\partial y_1} & \frac{\partial y_2}{\partial x_2} \cdot \frac{\partial E}{\partial y_2} & \cdots & \frac{\partial y_n}{\partial x_n} \cdot \frac{\partial E}{\partial y_n} \end{bmatrix}_{f\times n}$$

$$\frac{\partial y_1}{\partial x_1} = \begin{bmatrix} y_{11}(1-y_{11}) & -y_{12}y_{11} & \cdots & -y_{1f}y_{11} \\ -y_{11}y_{12} & y_{12}(1-y_{12}) & \cdots & -y_{1f}y_{12} \\ \vdots & \vdots & & \vdots \\ -y_{11}y_{1f} & -y_{12}y_{1f} & \cdots & y_{1f}(1-y_{1f}) \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1f} \\ y_{11} & y_{12} & \cdots & y_{1f} \\ \vdots & \vdots & \xleftrightarrow{f \text{ times}} & \vdots \\ y_{11} & y_{12} & \cdots & y_{1f} \end{bmatrix} \odot \begin{bmatrix} 1-y_{11} & -y_{11} & -y_{11} & \cdots & -y_{11} \\ -y_{12} & 1-y_{12} & -y_{12} & \cdots & -y_{12} \\ \vdots & & & & \\ -y_{1f} & -y_{1f} & \cdots & \cdots & 1-y_{1f} \end{bmatrix}$$

$$= M_1 \odot (I - M_1^T)$$

$\frac{\partial E}{\partial X} \to f\times n$

$\frac{\partial E}{\partial Y} \to f\times n$

$\boxed{\frac{\partial Y}{\partial X} \to n\times n ???!?}$ → Could not find solution to this.

I need to learn something about tensor contractions to be able to fully understand how this works.

$$e(x_i) = \sum_{k=1}^{j} (y - y$$

$X_i \sim \text{Binomial}(n,p)$, $n=3,\ p=\theta$

$$L(x_1,x_2,x_3,x_4;\theta) = \prod_{i=1}^{4} \binom{3}{x_i}\cdot \theta^{x_i}\cdot(1-\theta)^{3-x_i}$$

$$= \binom{3}{1}\binom{3}{3}\binom{3}{2}\binom{3}{2}\cdot \theta^{1+3+2+2}\cdot(1-\theta)^{3-(1+3+2+2)}$$

$X \sim \text{Binomial}(n,p)$

$$= 3\cdot 1\cdot 3\cdot 3\cdot \theta^{8}\cdot(1-\theta)^{-5}$$ → 12 not −5

$$p(X=x) = \binom{n}{x} p^{x}\cdot(1-p)^{n-x}$$

$$= \frac{27\,\theta^{8}}{(1-\theta)^{5}}$$

$$\frac{d}{d\theta}\left(L(X;\theta)\right) = 0$$

$$\frac{8\cdot\theta^{7}\cdot(1-\theta)^{5} + 5\,\theta^{8}(1-\theta)^{4}}{(1-\theta)^{10}} = 0$$

$$\theta^{7}\cdot(1-\theta)^{4}\left[8(1-\theta)+5\theta\right] = 0$$

$$\left[8-8\theta+5\theta\right] = 0$$

$$\theta^{7}(1-\theta)^{4}\left[8-3\theta\right]$$

$$\boxed{\theta = 0, 1, 3/8}$$

$(x_1, x_2, x_3, \ldots, x_n)$, $x_i \sim \text{Exponential}(\theta)$

$$L(X;\theta) = \prod_{i=1}^{n} \theta e^{-\theta x_i}$$

$$= \theta^{n}\cdot e^{-\theta\left(\sum_{i=1}^{n} x_i\right)}$$

$$\frac{d}{d\theta}\left(L(X;\theta)\right) = n\,\cancel{\theta^{n-1}}\ \cancel{e^{-\theta\left(\sum_{i=1}^{n}x_i\right)}} - \cancel{\theta^{n}}\ \theta\cdot\sum_{i=1}^{n}x_i\ \cancel{e^{-\theta\left(\sum_{i=1}^{n}x_i\right)}} = 0$$

$$= n - \theta\left(\sum_{i=1}^{n} x_i\right) = 0$$

$$\theta = \frac{n}{\sum_{i=1}^{n} x_i}$$

$$X_i \sim \text{Binomial}(m, \theta)$$

$$L(X;\theta) = \prod_{i=1}^{n} \binom{m}{x_i} \cdot \theta^{x_i} \cdot (1-\theta)^{m-x_i}$$

$$= \left[\prod_{i=1}^{n} \binom{m}{x_i}\right] \cdot \left(\frac{\theta}{1-\theta}\right)^{\sum x_i} \cdot (1-\theta)^{mn} = 0$$

$$\sum x_i \cdot \theta^{\sum x_i - 1} (1-\theta)^{mn - \sum x_i} - (mn - \sum x_i)(1-\theta)^{mn-\sum x_i - 1} \theta^{\sum x_i} = 0$$

$$\sum x_i \cdot \theta^{-1} - (mn - \sum x_i)(1-\theta)^{-1} = 0$$

$$S(1-\theta) - (mn - S)\theta = 0$$

$$S - \cancel{S\theta} - mn\theta + \cancel{S\theta} = 0$$

$$\boxed{\theta = \frac{S}{mn} = \frac{\sum_{i=1}^{n} x_i}{mn}}$$

$$\ell(\theta) = \sum_{i=1}^{n} \left(-\frac{x_i}{\theta} - \log\theta \right)$$

$$\frac{\partial \ell(\theta)}{\partial \theta} = 0 \Rightarrow \sum_{i=1}^{n} + \frac{x_i}{\theta^2} - \frac{1}{\theta} = 0$$

$$\sum_{i=1}^{n} (x_i - \theta) = 0$$

$$\sum_{i=1}^{n} x_i = n\theta \Rightarrow \boxed{\theta = \frac{\sum_{i=1}^{n} x_i}{n}}$$

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

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

$$\boxed{E(X) = 6}$$

$$\ell(\lambda) = \sum x_i \log\lambda - \lambda - \log x_i!$$

$$\ell'(\lambda) = \sum \frac{x_i}{\lambda} - 1 = 0$$

$$\frac{s}{\lambda} - n = 0$$

$$\boxed{\lambda = \frac{s}{n} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

$H(p, q)$ is convex in q

$$f(\vec{x}) = \begin{bmatrix} e^{x_1}/\Sigma \\ e^{x_2}/\Sigma \\ \vdots \\ e^{x_t}/\Sigma \end{bmatrix}$$

$$\hat{y}_i = \frac{e^{x_i}}{\sum_k e^{x_k}}$$

$$\frac{\partial \hat{y}_i}{\partial x_i} = \frac{e^{x_i} \sum e^{x_k} - e^{x_i} \cdot e^{x_i}}{\left(\sum_k e^{x_k}\right)^2}$$

$$= \left(\frac{e^{x_i}}{\sum_k e^{x_k}}\right)\left(1 - \frac{e^{x_i}}{\sum_k e^{x_k}}\right) = \hat{y}_i(1 - y_i)$$

$$\frac{d \hat{y}_i}{\partial x_j} = \frac{-e^{x_i} \cdot e^{x_j}}{\left(\sum e^{x_k}\right)^2} = -\hat{y}_i \hat{y}_j$$

$$L = -\sum y_i \ln \hat{y}_i$$

$$\frac{dL}{\partial \hat{y}_i} = -\sum_i \frac{y_i}{\hat{y}_i}$$

$$\frac{\partial L}{\partial x_j} = -\sum_{i \neq j} \frac{y_i}{\hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial x_j} - y_j \frac{1}{\hat{y}_j} \frac{\partial y_j}{\partial \hat{x}_j}$$

Tany → approximation

$H(p, q)$

$H(p,q) = -\sum_i p_i \log q_i$

$KL(p,q) = -\sum_i p_i \log \frac{q_i}{p_i}$

1 - 3hr

2 - 1hr @

3 - 2hr @

Probab of choosing any 1 way $= \frac{1}{3}$

$E[X] = \sum p \cdot x = \frac{1}{3} \times 3 + \frac{1}{3} \cdot 1 \cdot E[X] + \frac{1}{3} \cdot 2\, E[X]$

$= 1 + \frac{E[X]}{3} + \frac{2}{3} E[X]$

$E[X] = 1 + E[X]$

$p(x;\theta) = \begin{cases} \frac{1}{\pi\theta^2} ; & \|x\| \le \theta \\ 0 ; & \text{otherwise} \end{cases}$

$\boxed{\|x\| = \sqrt{x_1^2 + x_2^2}}$

pm

$\prod p(x;\theta)$

$= \boxed{\prod_{i=1}^{n} \frac{1}{\pi\theta^2}}$

$=$

$\text{sign}(x) = \begin{cases} +1 & x > 0 \\ 0 & x \le 0 \end{cases}$

$$\boxed{p(x;\theta) = \frac{1}{\pi\theta^2} \cdot \text{sign}(\theta - \|x\|)}$$

$\left(\frac{1}{\pi\theta^2}\right)^m$

$\boxed{x_1^2 + x_2^2 \le \theta^2}$

Consider them in m.

$-m \log \pi\theta^2$

$-\frac{m}{\pi\theta^2} \times 2\pi\theta = 0$

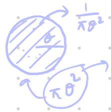
$$\frac{-m}{\pi\theta^2} \times 2\pi\theta = 0$$

$$\frac{-2m}{\theta} = 0$$

$$\theta \to \infty$$

$$\sqrt{x_1^2 + x_2^2} = \|x\|$$

$$x_1^2 + x_2^2 \le \theta^2$$



$$p(x;\theta) = \begin{cases} \frac{1}{\pi\theta^2} & \|x\| \le \theta \\ 0 & \text{otherwise} \end{cases}$$

$$\approx \boxed{\pi\theta^2}$$

$$\left(\frac{1}{\pi\theta^2}\right)^{\pi\theta^2}$$

$$= \pi\theta^2 \log(\pi\theta^2)$$

$$-2\pi\theta \log(\pi\theta^2) - \frac{\pi\theta^2}{\pi\theta^2} \times 2\pi\theta =$$

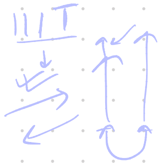
$$-2\pi\theta\left(1 + \log(\pi\theta^2)\right) = 0$$

$$\log(\pi\theta^2) = -1$$

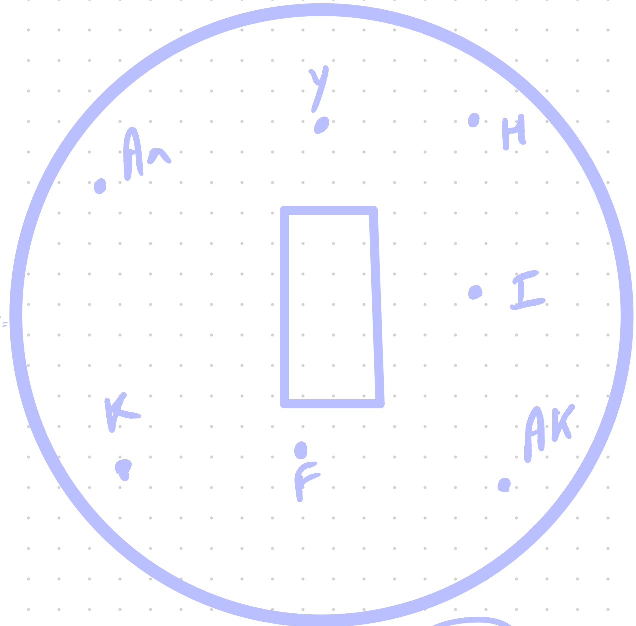
$$\pi\theta^2 = \frac{1}{e}$$

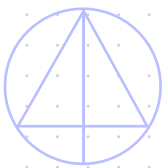
$$\theta^2 = \frac{1}{e\pi}$$

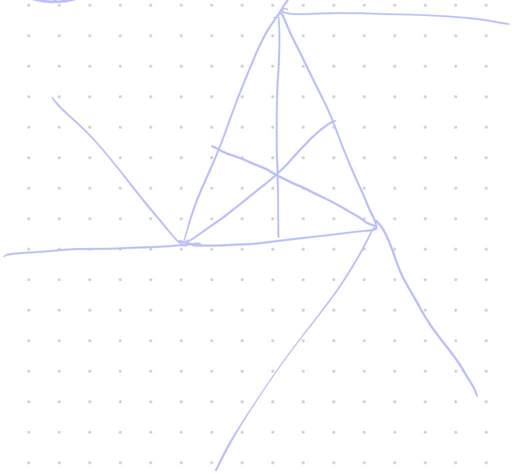
$$\boxed{\theta = \frac{1}{\sqrt{e\pi}}}$$



$$\boxed{\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}}$$

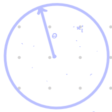






GR

$\boxed{|x_i| \le \theta}$



$\left(\frac{1}{\pi\theta^2}\right)^n = L(\theta)$

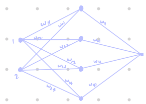
$\theta = \max(x) \in \{x_i\}$ $\quad \hat{\theta} = x_{max}$

$P(x_1) \cdot P(x_2) \cdots P(x_n)$

$\theta_{MLE} = \frac{\sum x_i}{n}$ $\quad \max(x_i)$

$\frac{n}{\sum x_i}$

Row vector Gradient



$$X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix}_{3\times 2}$$

$$W^{(1)} = \begin{bmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \end{bmatrix}_{2\times 4} \qquad W^{(2)} = \begin{bmatrix} w_{11} \\ w_{21} \\ w_{31} \\ w_{41} \end{bmatrix}_{4\times 1}$$

$$\boxed{Z^{(2)} = XW^{(1)}} \qquad \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix} \begin{bmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \end{bmatrix}$$

$$Z^{(2)} = \begin{bmatrix} w_{11}x_{11}+w_{21}x_{12} & w_{12}x_{11}+w_{22}x_{12} & w_{13}x_{11}+w_{23}x_{12} & w_{14}x_{11}+w_{24}x_{12} \\ \hline \\ \end{bmatrix}_{3\times 4} \longrightarrow Z_i^{(2)}$$

$$a^{(2)} = f(Z^{(2)}) \; ; \; f(x) = \frac{1}{1+e^{-x}}$$

$$Z^{(3)} = a^{(2)}W^{(2)} = \begin{bmatrix} a_{11}^{(2)} & a_{12}^{(2)} & a_{13}^{(2)} & a_{14}^{(2)} \\ a_{21}^{(2)} & a_{22}^{(2)} & a_{23}^{(2)} & a_{24}^{(2)} \\ a_{31}^{(2)} & a_{32}^{(2)} & a_{33}^{(2)} & a_{34}^{(2)} \end{bmatrix}_{3\times 4} \begin{bmatrix} w_{11} \\ w_{21} \\ w_{31} \\ w_{41} \end{bmatrix}_{4\times 1}$$

$$= \begin{bmatrix} a_{11}^{(2)}w_{11} + a_{12}^{(2)}w_{21} + a_{13}^{(2)}w_{31} + a_{14}^{(2)}w_{41} \\ \hline \\ \end{bmatrix}_{3\times 1} \longrightarrow \text{first example}$$

$$\hat{y} = f(Z^{(3)})_{3\times 1} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \end{bmatrix}$$

$$J = \frac{1}{2}\sum_{i=1}^{3}(y_i - \hat{y}_i)^2$$

$$\frac{\partial J}{\partial w^{(2)}} = \sum_{i=1}^{3}(y_i - \hat{y}_i) \cdot - \frac{\partial \hat{y}_i}{\partial w^{(2)}}$$

$$= -\sum_{i=1}^{3}(y_i - \hat{y}_i) \; \frac{\partial \hat{y}_i}{\partial z_i^{(3)}} \cdot \frac{\partial z_i^{(3)}}{\partial w^{(2)}}$$

$$\frac{\partial J}{\partial w^{(2)}} = -\sum_{i=1}^{3} \underbrace{(y_i - \hat{y}_i) \cdot \sigma'(z_i^{(3)})} \cdot a_i^{(2)}$$

$$\hat{y}_i = \sigma(z_i^{(3)})$$

$$z_i^{(3)} = \underset{1\times 4}{a_i^{(2)}} \; \underset{4\times 1}{W^{(2)}}$$

$$z_i^{(3)} = \begin{bmatrix} a_{11}^{(2)} & a_{12}^{(2)} & a_{13}^{(2)} & a_{14}^{(2)} \end{bmatrix} \begin{bmatrix} w_{11} \\ w_{21} \\ w_{31} \\ w_{41} \end{bmatrix}$$

$$z_i^{(3)} = a_{11}^{(2)}w_{11} + a_{12}^{(2)}w_{21} + a_{13}^{(2)}w_{31} + a_{14}^{(2)}w_{41}$$

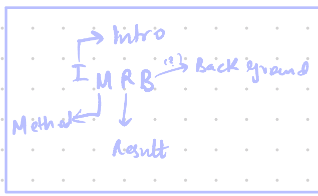
$$\frac{\partial z_i^{(3)}}{\partial w_{41}} = a_{14}^{(2)}$$

$$\frac{\partial z_i^{(3)}}{\partial w^{(2)}} = \begin{bmatrix} \frac{\partial}{\partial w_{11}} z_i^{(3)} & \frac{\partial z_i^{(3)}}{\partial w_{21}} & \frac{\partial z_i^{(3)}}{\partial w_{31}} & \frac{\partial z_i^{(3)}}{\partial w_{41}} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}^{(2)} & a_{12}^{(2)} & a_{13}^{(2)} & a_{14}^{(2)} \end{bmatrix}$$

$$= a_i^{(2)}$$

Informative



End Semester Prep

SMAI Final Exam - syllabus and instructions
by Ravi Kiran Sarvadevabhatla - Sunday, 12 November 2023, 12:44 AM

SMAI final exam will contain questions from following topics:

- Basics of Probability
- Decision Trees
- k-NN
- Linear Regression
- Naive Bayes
- PCA
- Logistic Regression
- Neural Networks (MLP, CNN)
- Bias-Variance, Regularization
- Ensemble Methods
- Kernel Density Estimation
- HMM

The 3 hour exam will have True/False, Multiple Choice, Long Answer questions. 2 or 3 questions will be on proofs. Others will be descriptive, factual, numerical questions. Numerical questions will not require a calculator.

How to prepare: For each topic, try initially writing the key definitions and equations from memory. Revise the slides based on how well you were able to recall the definitions/equations. Preparing a reference sheet (mentioned below) may also help indirectly revise the exam syllabus.