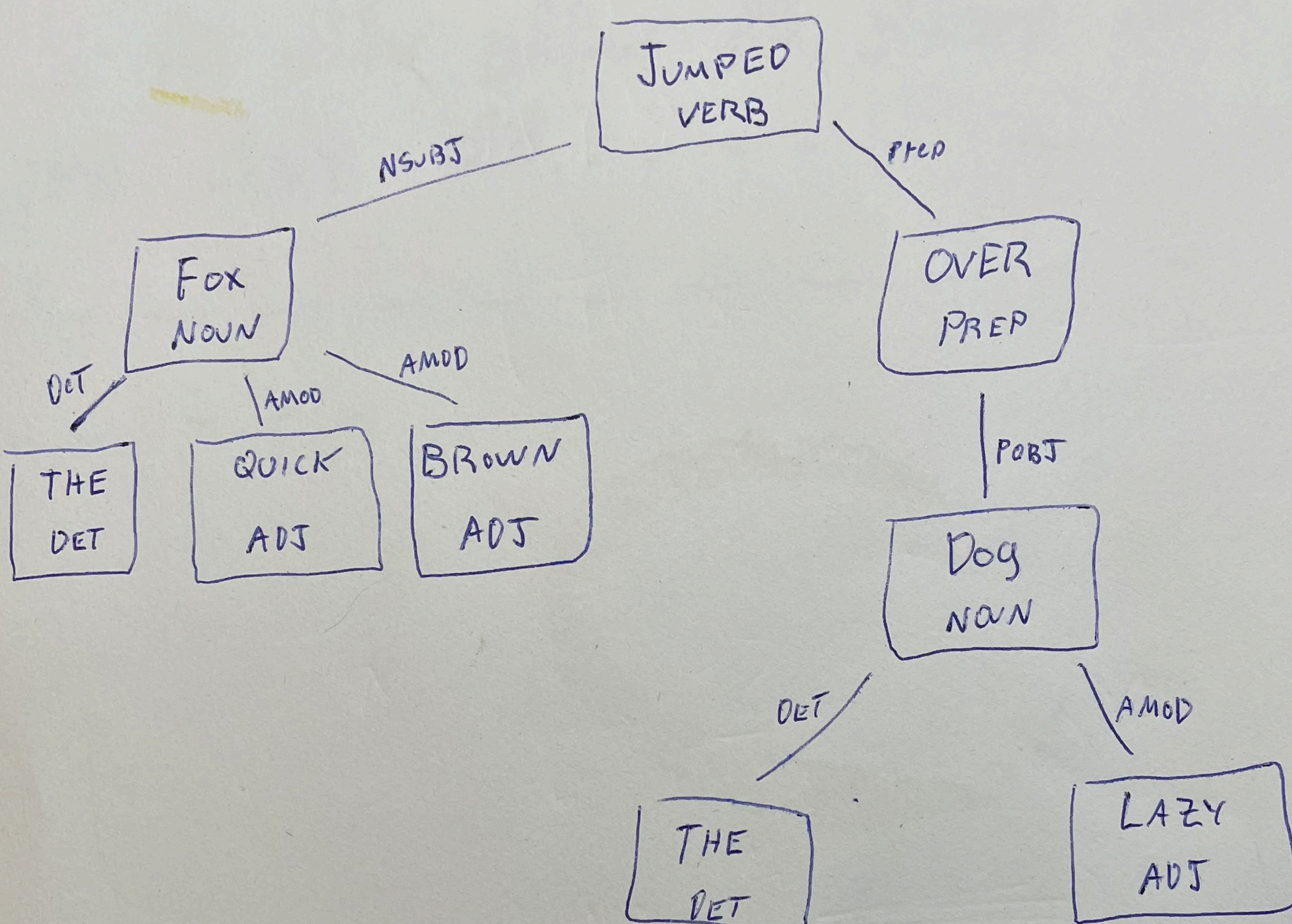
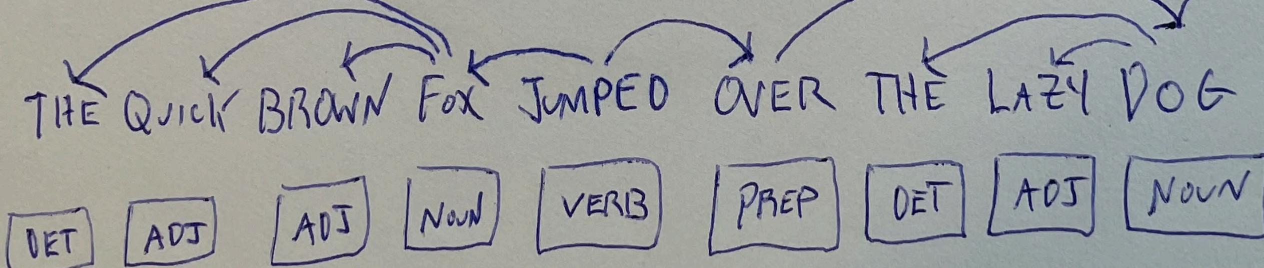
DEPENDENCY PARSING

TEXT : "THE QUIK BROWN FOX JUMPED OVER THE LAZY DOG"





THE ML-APPROACH TO QA

THE QUICK BROWN FOX JUMPED OVER THE LAZY DOG.

FEATURE REPRESENTATION

$$X \in \mathbb{R}^{M \times N}$$

M = NUM OF OBSERVATIONS

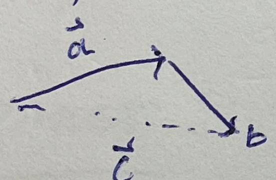
N = NUM OF FEATURES

$$X = \begin{bmatrix} X_{11} & \cdots & X_{1N} \\ \vdots & \ddots & \vdots \\ X_{M1} & \cdots & X_{MN} \end{bmatrix}$$

NORM

$f(\cdot)$ MUST SATISFY:

① $f(\vec{x}) = 0$ iff $\vec{x} = \vec{0}$

② 

$a + b \geq c$

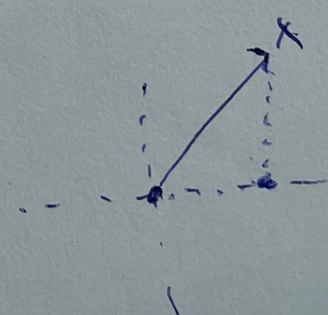
$f(\vec{a} + \vec{b}) \leq f(\vec{a}) + f(\vec{b})$

③ $f(\alpha \vec{x}) = |\alpha| f(\vec{x})$

L_p NORM: $\|x\|_p = \left(\sum_{i=1}^{N} |x_i|^p \right)^{\frac{1}{p}}$

FROBENIUS NORM: $\|A\|_F = \sqrt{\sum_{i,j} A_{ij}^2}$

$L_2 \rightarrow \|x\|_2 = \sqrt{\sum_{i=1}^{N} x_i^2}$

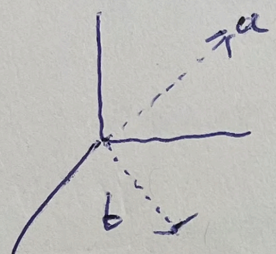


$L_1 \rightarrow \|x\|_1 = \sum_{i=1}^{N} |x_i|$

DISTANCES

$$\|x_1 - x_2\|_2 = \sqrt{\sum_{i=1}^{N} \left(x_i^{(1)} - x_i^{(2)}\right)^2}$$

$$\|x_1 - x_2\|_1 = \sum_{i=1}^{N} \left|x_i^{(1)} - x_i^{(2)}\right|$$



INNER PRODUCT:

$$\langle a, b \rangle = a \cdot b = \sum_{i=1}^{N} a_i b_i \in \mathbb{R}$$

OUTER PRODUCT:

$$\vec{a} \otimes \vec{b} = \vec{a}\vec{b}^T = \begin{bmatrix} a_1 b_1 & \cdots & a_1 b_N \\ \vdots & \ddots & \vdots \\ a_N b_1 & \cdots & a_N b_N \end{bmatrix}$$

$$\cancel{\vec{a} \otimes \vec{b} = \mathrm{Tr}(}$$

$$\vec{a} \cdot \vec{b} = \mathrm{Tr}(\vec{a}\vec{b}^T)$$

SVD

$X \in \mathbb{R}^{M \times N}$

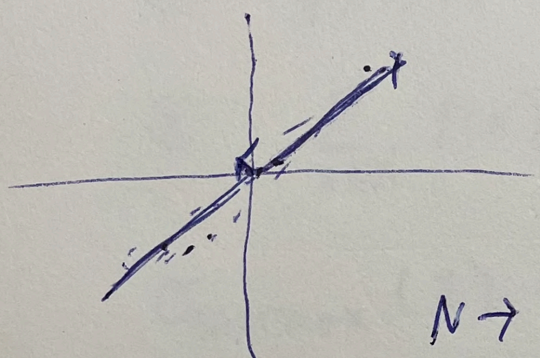
U, V ARE ORTHOGONAL

$$\underset{M \times N}{X} = \underset{M \times M}{U}\ \underset{M \times N}{\Sigma}\ \underset{N \times N}{V^T}$$

$U^T = U^{-1}$

$V^T = V^{-1}$

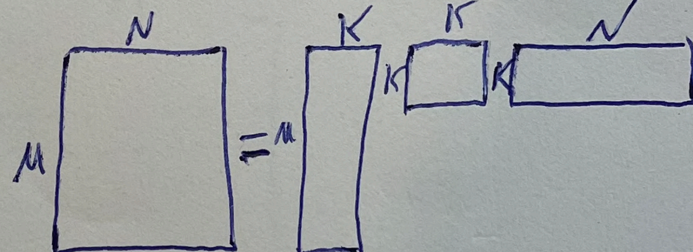
$$\Sigma = \begin{array}{c} N \\ \\ M-N \end{array} \left[\begin{array}{cccc} \sigma_1 & \sigma_2 & \cdots & \sigma_N \\ \hline & \text{NULL SPACE} & & \end{array}\right]$$



$N \to K \quad K \ll N$

$\text{RANK}(X) = \text{NUM NON-ZERO } \sigma_i\text{'s}$

$$\underset{M \times K}{X = U}\ \underset{K \times K}{\Sigma}\ \underset{K \times N}{V^T}$$

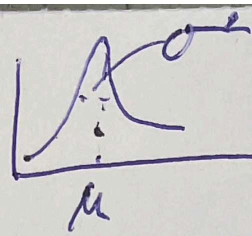


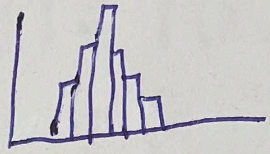
$\underset{1 \times N}{X'} \to \underset{1 \times K}{\hat{X}'}$

$$\boxed{\underset{1 \times K}{\hat{X}'} = \underset{1 \times N}{X'}\ \underset{N \times K}{V}}$$

PROJECTION OF X' FROM $\mathbb{R}^N$ TO $\mathbb{R}^K$ WHERE $K < N$.

- $\text{CONDITION}(X) = \dfrac{\sigma_1}{\sigma_N}$

- PDF → $P(X)$ 

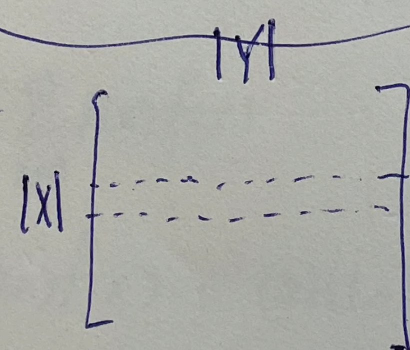
- PMF → $P(X)$ 

PMF MUST SATISFY:

① $\forall x \in X \quad 0 \leq P(x) \leq 1$

② $\sum_{x \in X} P(x) = 1$

$$\sigma_{\text{SOFTMAX}}(X) = \frac{e^{x}}{\sum_{x' \in X} e^{x'}}$$

JOINT: $P(x,y) =$ 

$x \in X$

$y \in Y$

MARGINAL DIST: $P(x) = \sum_{y \in Y} P(x,y)$

CONDITIONAL DIST: $P(y|x) = \dfrac{P(x,y)}{P(x)}$

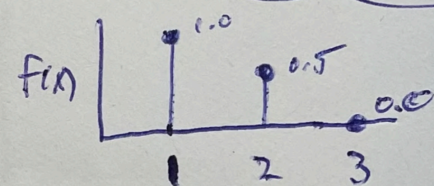
PRODUCT RULE:

$$P(x,y) = P(x)\cdot P(y\mid x)$$
$$= P(y)\cdot P(x\mid y)$$

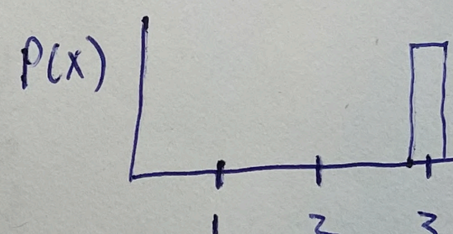
INDEPENDENCE: $P(x,y) = P(x)\cdot P(y)$

CONDITIONAL INDEPENDENCE: $P(x,y\mid z) = P(x\mid z)\,P(y\mid z)$

EXPECTATION: $f(x)$



$$E[f(x)] = \frac{1}{|X|}\sum_{x\in X} f(x) = \underline{1.5}$$



$$E_{x\sim p}[f(x)] = \sum_{x\sim p} p(x)f(x)$$

$$= \sum_{x\sim p} p(x)f(x) = (0)(1.0) + (0)(0.5) + (1)(0)$$

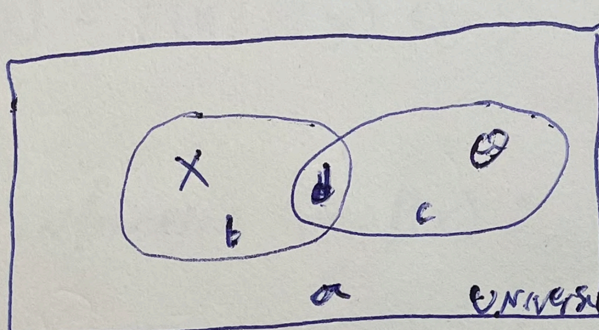
$$= \underline{0}$$

BAYES RULE

X, θ

$$P(X,\theta) = P(X|\theta)P(\theta)$$
$$= P(\theta|X)P(X)$$

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}$$



$P(\theta) = \frac{c}{a}$ $\quad$ $P(\theta|X) = \frac{d}{b}$

$P(X) = \frac{b}{a}$ $\quad$ $P(X|\theta) = \frac{d}{c}$

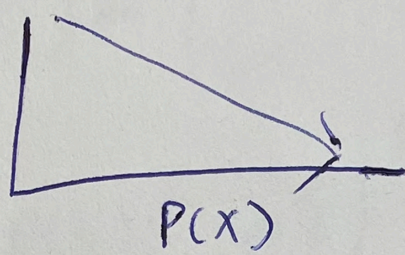
$$\frac{d}{\cancel{c}} \cdot \frac{\cancel{c}}{a} = \frac{d}{\cancel{b}} \cdot \frac{\cancel{b}}{a} = \frac{d}{a}$$

$$\text{POSTERIOR} = \frac{\text{Likelihood} \cdot \text{Prior}}{\text{EVIDENCE}}$$

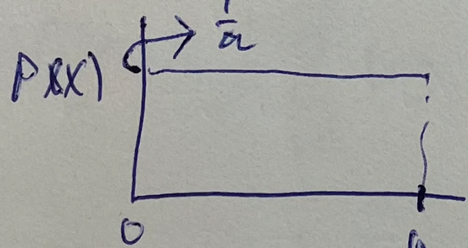
Info Theory

$P(x)$ + observe $x = x_0$

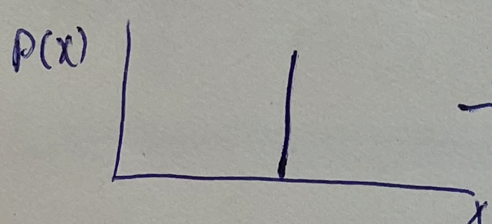
- $P(x = x_0) = 1 \rightarrow x_0 + I(x = x_0) = 0$
- $I(x)$



- $P(x, y) = P(x)P(y) \iff I(x, y) = I(x) + I(y)$
- Self information of $P(x)$: $I(x) = -\log P(x)$
- $H(P) = E_{x \sim p}[I(x)] = -\sum_{x \sim p} P(x) \log P(x)$
-

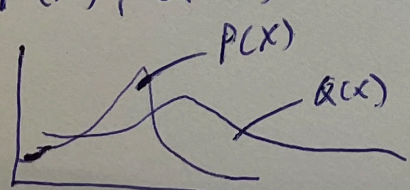


$\rightarrow H$ is Big



$\rightarrow H = 0$

- $P(x), Q(x)$



$\cancel{D_{KL}(P \| Q) = E_{x \sim p}\left[P(x) \right.}$

$D_{KL}(P \| Q) = E_{x \sim p}\left[\log \frac{P(x)}{Q(x)} \right]$

Maximum Likelihood

$P(X,Y) = P(X)P(Y)$

L1-11

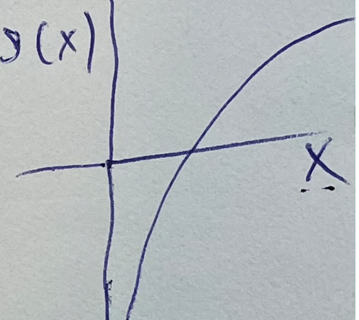
$$P(y_i \mid x_i ; \Theta)$$

$$P(Y \mid X ; \Theta) = \left(\prod_{i=1}^{M} P(y_i \mid x_i ; \theta) \right)^{\frac{1}{M}}$$

$$L \rightarrow \underbrace{\left(\prod_{i=1}^{M} P(y_i \mid x_i ; \theta) \right)^{\frac{1}{M}}}_{\text{Big No No!}}$$

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmax}} \; \frac{1}{M} \sum_{i=1}^{M} \log P(y_i \mid x_i ; \theta)$$

$\log(x)$



$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmax}} \; \frac{1}{M} \sum_{i=1}^{M} \log P(y_i \mid x_i \; \theta)$$

$$= \underset{\Theta}{\operatorname{argmin}} \; \frac{1}{M} \sum_{i=1}^{M} - \log P(y_i \mid x_i ; \Theta)$$

— PARAMETERIZAT.

$P(Y \mid X ; \theta) \Leftrightarrow Q(X)$

LABEL DISTRIBUTION $\Leftrightarrow P(X)$