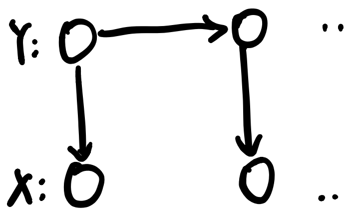
HMM



Forward

$\alpha_{t,k} = P(x_1, x_2, \cdots x_t, y_t^k = 1)$

$\alpha_{t,k} = \sum_i P(x_1, x_2 \cdots x_{t-1}, y_{t-1}^i = 1) \cdot P(y_t^k \mid y_{t-1}^i = 1) \cdot P(x_t \mid y_t^k = 1)$

$= P(x_t \mid y_t^k = 1) \sum_i P(x_1, x_2, \cdots x_{t-1}, y_{t-1}^i = 1) P(y_t^k \mid y_{t-1}^i = 1)$

$\Rightarrow \alpha_t^k = P(x_t \mid y_t^k = 1) \cdot \sum_i \alpha_{t-1,i}\, a_{i,k}$

$P(x) = \sum_i \alpha_t^i$ ✓

Backward

$\beta_{t,k} = P(x_{t+1}, x_{t+2} \cdots x_T \mid y_t^k = 1)$

$\beta_{t,k} = \sum_i P(x_t, x_{t+1} \cdots x_T \mid y_{t-1}^i = 1) \cdot P(y_t^k \mid y_{t-1}^i) \cdot P(x_t \mid y_t^k)$

$= \sum_i \beta_{t-1,i}\, a_{i,k}\, P(x_t, y_t^k)$

$P(x) = \sum_k P(x_1, x_2 \cdots x_T \mid y_1^k = 1) \cdot P(y_1^k = 1)$ Prior $p(y_t^i = 1) = \pi_i$

$= \sum_k \beta_{1,k}\, P(x_1 \mid y_t^k = 1) \cdot \pi_i = \sum_i \beta_{1,i} \boxed{\pi_i}\, b_{i,1}$

Forward - Backward

$P(O \mid I) = \frac{P(x, Y)}{P(x)} = \frac{\alpha_{t,i} * \beta_{t,i}}{P(x)}$ ~ Posterior

Verterbi : Finding $y = y_1, y_2 \cdots y_T$ s.t. $P(y \mid x) \uparrow$

$y^* = \arg\max_y P(y \mid x)$

$V_t^k = \max P(y_1, y_2 \cdots y_{t-1}, x_1, x_2 \cdots x_{t-1}, x_t, y_t^k = 1)$

Recur: $V_t^k = P(x_t \mid y_t^k = 1) \max_i a_{i,k} V_{t-1}^i$

~ Actually Dynamic Programming

EM Algorithm

Original: $\ell(\theta) = \sum_{i=1}^{m} \log p(x_i;\theta)$

$\Rightarrow$ Hidden: $\ell(\theta) = \sum_{i=1}^{m} \log P(x_i;\theta) = \sum_{i=1}^{m} \log \sum_{z} p(x_i, z; \theta)$

Main idea: Considering $p(z_i)$

① Random Model: $p(z_i)$ GET!

② $L(\theta) = \prod \log \sum_z p(x_i, z, \theta) \Rightarrow$ New model θ } Recursive

Q: Jensen ~ $\log \sum_z p(x_i, z; \theta) = \log \sum_z Q(z) \boxed{\frac{p(x_i, z; \theta)}{Q(z)}}$ Assigned Y

$$= \log \sum_z Q(z) Y = \log \sum_z P(Y)\cdot Y = \log E(Y)$$

$$\geq E(\log Y) = \sum P(Y) \log Y = \sum_z Q(z) \log \frac{p(x_i, z; \theta)}{Q(z)}$$ Lower Bound

$$\ell(\theta) = \sum_{i=1}^{m} \log \sum_z P(x_i, z; \theta) \geq \sum_i \sum_z Q(z) \log \frac{P(x_i, z; \theta)}{Q(z)}$$

To make it TIGHT!

$$Y = \frac{p(x_i, z; \theta)}{Q(z)} = C$$ ~ According to Jensen

$$\sum_z Q(z) = 1 = \frac{\sum_z P(x_i, z; \theta)}{C} = 1$$

$$Q(z) = P(x_i, z; \theta) / C = \frac{p(x_i, z; \theta)}{\sum_z p(x_i, z; \theta)} = p(z_i \mid x_i; \theta)$$

{ E-step: According to θ. $\Rightarrow P(z_i \mid x_i; \theta) = Q$

{ M-step: Max $L(\theta) \Rightarrow$ Getting new θ!

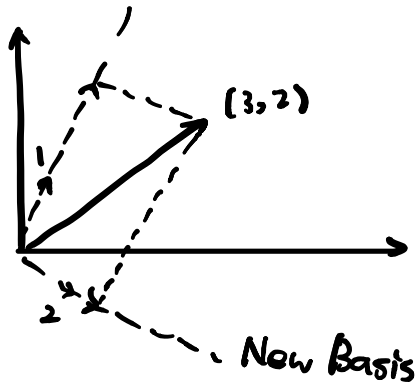
Math of PCA

$\langle \vec{a}, \vec{1} \rangle = L[\mathrm{Proj}_1(a)]$

$\vec{a} = (3, 2)$ ~ The weight of basis!

内积$\langle \vec{a}, \vec{b} \rangle$：$\vec{a}$在$\vec{b}$基上的分量

$A b_1 = b_2$ where A: Basis Transformation



$$\begin{bmatrix} \text{Basis } 1 \\ \text{Basis } 2 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} \text{New} \\ \text{Basis} \\ (3,2) \end{bmatrix}$$

Generally

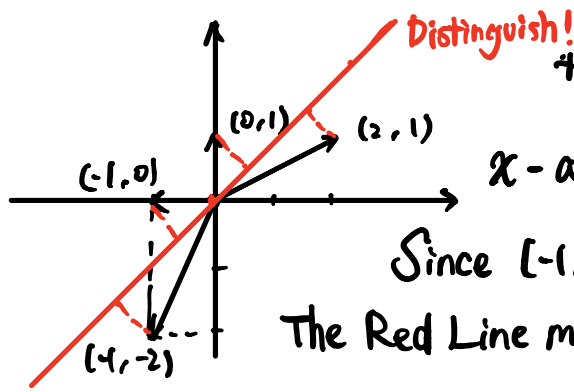
$$\underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{bmatrix}}_{\text{Basis of } R^N} \overbrace{\begin{bmatrix} a_1 & a_2 & \cdots & a_M \end{bmatrix}}^{M \text{ vector to be transformed}} = \underbrace{\begin{bmatrix} b_1 a_1 & \cdots & b_1 a_M \\ b_2 a_1 & \cdots & b_2 a_M \\ \vdots & & \vdots \\ b_N a_1 & \cdots & b_N a_M \end{bmatrix}}_{\text{新空间中的“坐标”}}$$

✰ How to choose BASIS

Or say: N-D vector $\Rightarrow$ K-D. How should we choose K basis!

e.g. $\begin{bmatrix} -1 & -1 & 0 & 2 & 0 \\ -2 & 0 & 0 & 1 & 1 \end{bmatrix}$ How to find 1-D Interpretation

A Simple Intuition is to draw a line that can distinguish all the data vec.



x-axis is not a good choice

Since $(-1, 0)$, $(-1, -2)$ will be the same

The Red Line may be better

Pure Mathmetical Interpretation.

α. Varience : Distinguishable $\Longleftrightarrow$ max $Var(a) = \frac{1}{m}\sum_i (a_i - \mu)^2$

Since $\mu = 0 \Rightarrow Var(a) = \frac{1}{m}\sum_i a_i^2$

<归一化> Problem Reduce to find 1-D basis s.t. max $Var(a)$

β. Covarience Σ

High dimension case : ① each dimension : Distinguishable <same as 1-D>

② No Similar feature selections!

选取的划分平面不应重合! $\Rightarrow$ Covarience

$Cov(a,b) = \frac{1}{m}\sum_i a_i b_i$ ~ 归一化了!

Objective : Select K-dim basis from N-dim

s.t. max $Var(a)$. min $Cov(a,b)$

• Now consider $X = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}$ $(n \times m)$

Varience

$$\frac{1}{m}XX^T = \frac{1}{m}\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}\begin{bmatrix} a_{11} & \cdots & a_{n1} \\ a_{12} & \cdots & a_{n2} \\ \vdots & & \vdots \\ a_{1m} & \cdots & a_{nm} \end{bmatrix} = \begin{bmatrix} \frac{1}{m}\sum_i a_{1i}^2 & \cdots \\ \frac{1}{m}\sum_i a_{1i}a_{2i} & \cdots \\ \vdots & \end{bmatrix}$$

Covarience

Now the objective become

If we can find Linear Transformation P.

s.t.

Di... al Matrix

$$\frac{1}{m} PX(PX)^T = D \leftarrow \text{Diagonal Matrix.}$$

$$= \frac{1}{m} PXX^T P^T \Rightarrow P\text{'s rows are the trans.}$$

Steps ① Normalization

② $C = \frac{1}{m} XX^T$ ~ Covarience Matrix

③ Getting λ & eigenvector

④ $Y = PX$