$x_1: x_{11}, x_{12}, \cdots x_{1D}$

$x_2: x_{21}, x_{22}, \cdots x_{2D}$

$\vdots$

$x_{N1}, x_{N2}, \cdots, x_{ND}$

features

$\overline{x}$ $\overline{x^{(1)}}$ $\overline{x^{(2)}}$

$x$ $\overline{x_1}$ $\overline{x_2}$ → for now consider this feature

$\mathbb{R}^{10,000}$ if $D = 10,000$

$\mathbb{R}^{40}$

First component → combination of all 10,000

$\vdots$

$40^{th}$ component → " " " " "

$f_1$ $f_2$ - - - - - $f_{100}$

$D_1$

$D_2$

$\vdots$

$D_{100\,000}$

→ Variability of data : $\Sigma$ variability of all features

$$= \frac{1}{N}\left(\sum_{i=1}^{N}(x_{i1}-\overline{x_1})^2 + \sum_{i=1}^{N}(x_{i2}-\overline{x_2})^2 + \sum_{i=1}^{N}(x_{iD}-\overline{x_D})^2\right)$$

$$\rightarrow \frac{1}{N}\sum_{j=1}^{D}\sum_{i=1}^{N}(x_{ij}-\overline{x_j})^2$$

$D_1$

$\vdots$

$D_N$

$\underbrace{f_1 \; \cdots \; f_{100}}$

variability should be the same as that of 10,000

$$x_1 = \begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1D} \end{bmatrix}^T \quad x_2 = \begin{bmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{2D} \end{bmatrix}^T \cdots$$

$$(x_n - \bar{x})(x_n - \bar{x})^T$$

$$\begin{bmatrix} x_{n1} - \bar{x}_1 \\ x_{n2} - \bar{x}_2 \\ \vdots \\ x_{nD} - \bar{x}_D \end{bmatrix} \begin{bmatrix} x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 & \cdots \end{bmatrix}$$

$$\frac{1}{N}\sum_{n=1}^{N} (x_n - \bar{x})(x_n - \bar{x})^T$$

$$\begin{bmatrix} (x_{n1} - \bar{x}_1)^2 & (x_{n1} - \bar{x}_1)(x_{n2} - \bar{x}_2) & \cdots \\ (x_{n2} - \bar{x}_2)(x_{n1} - \bar{x}_1) & (x_{n2} - \bar{x}_2)^2 & \ddots \\ (x_{nD} - \bar{x}_D)(x_{n2} - \bar{x}_2) & & (x_{nD} - \bar{x}_D) \end{bmatrix}$$

$$\frac{1}{N}\begin{bmatrix} \sum_{n=1}^{N}(x_{n1} - \bar{x}_1)^2 & \sum_{n=1}^{N}(x_{n1} - \bar{x}_1)(x_{n2} - \bar{x}_2) & \sum_{n=1}^{N}(x_{n1} - \bar{x}_1)(x_{nD} - \bar{x}_D) \\ \sum_{n=1}^{N}(x_{n2} - \bar{x}_2)(x_{n1} - \bar{x}_1) & \sum_{n=1}^{N}(x_{n2} - \bar{x}_2) & \sum_{n=1}^{N}(x_{n2} - \bar{x}_2)(x_{nD} - \bar{x}_D) \\ \sum_{n=1}^{N}(x_{nD} - \bar{x}_D)(x_{n1} - \bar{x}_1) & \sum_{n=1}^{N}(x_{nD} - \bar{x}_D)(x_{n2} - \bar{x}_2) & \sum_{n=1}^{N}(x_{nD} - \bar{x}_D)^2 \end{bmatrix}$$

$= S$, where $S$ is the covariance matrix

Eigenvector

$$\begin{bmatrix} 3 & 1 & 2 \\ 4 & -1 & 3 \\ 3 & -5 & 4 \\ -1 & -2 & 3 \end{bmatrix}_{4\times 3} \begin{bmatrix} 2 \\ 3 \\ \\ \end{bmatrix}_{3\times 1} = \begin{bmatrix} \overset{\rightarrow LC}{\sim} \\ \\ \\ \\ \end{bmatrix}_{4\times 1}$$

LT: $\mathbb{R}^3 \longrightarrow \mathbb{R}^4$

So now you know where we're going

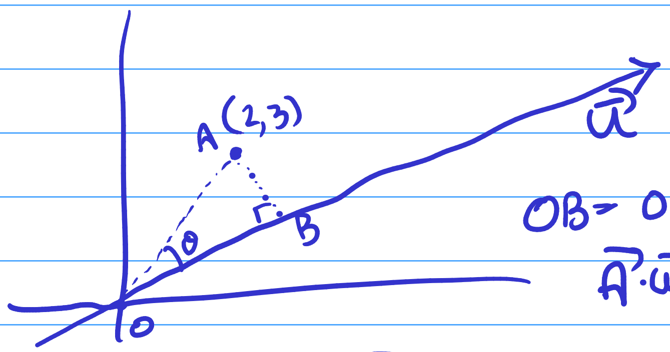
$$\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}_{10\times 1000} \begin{bmatrix} \\ \\ \\ \end{bmatrix}_{1000\times 1} = \begin{bmatrix} \\ \\ \\ \end{bmatrix}_{10\times 1}$$

Coming back to Eigen vectors

Lets say that by multiplying a square matrix on a 'crazy' vector $v$, we just get a scaled version of $v$ with no change in direction

$$A[v] = \lambda[v]$$

$v$ is the eigenvector




$OB = OA\cos\theta$

$\vec{A}\cdot\vec{u} =$ x coordinate along

$$u^T x_1 = \vec{u}\cdot\vec{x}$$

$$u^T X_{D\times N}$$

$$\frac{u^T x_1 + u^T x_2 \dots u^T x_N}{N}$$

$$= \quad u^T \frac{1}{N}\{x_1 + x_2 + x_3 \dots x_N\} = u^T \bar{x}$$

mean of projection = projection of mean

Variance $\Rightarrow \sum_{n=1}^{N} \frac{1}{N}\left(u^T(x_n - \bar{x})\right)^2 = \frac{1}{N}\sum_{n=1}^{N}\left((u^T(x_n-\bar{x})) \cdot (u^T(x_n-\bar{x}))^T\right)$

$$= \frac{1}{N}\sum_{n=1}^{N}\left(u^T(x_n-\bar{x})(x_n-\bar{x})^T u\right)$$

$$= \frac{1}{N}\left(u^T\left(\sum_{n=1}^{N}(x_n-\bar{x})(x_n-\bar{x})^T\right)u\right)$$

$$= u^T\left[\frac{1}{N}\left(\sum_{n=1}^{N}(x_n-\bar{x})(x_n-\bar{x})^T\right)\right]u$$

$$= u^T S u$$

now we need to maximize $u^T S u$

such that $u^T u = 1$

(u is a unit vector)

$$\max\left(u^T S u + \lambda(1 - u^T u)\right)$$

$$\nabla_u\left(u^T S u + \lambda(1-u^T u)\right) = 0$$

$$Su - \lambda u = 0$$

$\Rightarrow \quad Su = \lambda u$

Eigenvalue problem

$\Rightarrow$ these unitvectors maximise after the N data points are projected onto them

$u_1 \quad u_2 \quad \cdots \quad u_D$

$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_D$

what is the $u_i$ that maximises the variance

variance : $u_1^T S u_1 \quad \cdots \quad u_2^T S u_2 \quad \cdots \quad u_D^T S u_D$

Class went on for 2+hours jeez

Stopped here, pls watch recording

$S u_1 = \lambda u_1$

$u_1^T S u_1 = u_1^T \lambda u_1$

$u_1^T S u_1 = \lambda (u_1^T u_1) = \lambda(1)$

$u_1^T S u_1 = \lambda_1$

% variance

$$\frac{\lambda_1}{\text{Tot var}} \times 100$$

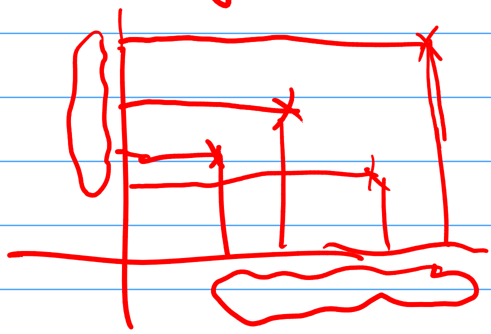
total variance = $Tr(S)$

$= \lambda_1 + \lambda_2 + \lambda_3 \cdots \lambda_D$ Sum of Eigenvalues

$$= \frac{\lambda_1}{\lambda_1 + \lambda_2 \cdots + \lambda_D} \times 100$$

$\vec{u}_1 \cdot \vec{u}_2 = 0$ too, you want $\perp$ axes

& linearly independent



1st vector
2nd vector

$$= \frac{\lambda_1 + \lambda_2}{[\quad]} \times 100$$

Like that get 10 vectors

$$\frac{\lambda_1 + \lambda_2 + \cdots + \lambda_{10}}{\quad} \times 100$$

10 dimensional vector

So if you want to reduce dimensionality put variance $\geq 95\%$

keep increasing vectors

END

$\{0, 1, 2 \cdots N\}$

Maximize $-p(x)\log(p(x))$

Subject to $x\,p(x) = \mu$

$(x-\mu)^2 p(x) = \sigma^2$

$$\max \left( \sum p(x)\log(p(x)) + \lambda\left(\sum_x p(x) - 1\right) \right)$$

$$\frac{\partial H}{\partial p} = 0 = \sum \log p \cancel{\$} - 1 + 0$$

$$\sum \log p = -1 \qquad p_i \geq 0$$

$$p = e^{-1}$$

$$-\sum p(x)\log(p(x)) + \lambda\left(\sum p(x) - 1\right)$$

$$\max\left( -\sum p(x)\log(p(x)) + \lambda\left(\sum p(x) - 1\right)\right)$$

$$\max \sum_{x_i} \left(-p(x)\log(p(x)) + \lambda p(x) - \lambda\right)$$

$$\cancel{p} - \log p(x) \cancel{\mp} 1 + \lambda = 0$$

$$p(x) = e^{\cancel{-p\ldots}\,\lambda - 1}$$

$$\forall p(x) \quad p(x) = e^{\lambda-1}$$

$$\sum p(x) = 1$$

$$\Rightarrow p(x) = \frac{1}{n} \qquad \lambda = -\log n + 1$$

$$H \quad -\sum p(x)\log p(x) = \left(\frac{1}{n}\log\frac{1}{n}\right) n = -\log n$$

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

$$-\int_D \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot \left(\log\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right) - \frac{(x-\mu)^2}{2\sigma^2}\right) dx$$

$$-\left[\int \left(\frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{(x-\mu)^2}{2\sigma^2}} \log\frac{1}{\sqrt{2\pi}\,\sigma}\right) dx - \int_{-\infty}^{\infty} \frac{(x-\mu)^2}{2\sigma^2} \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\,\sigma} dx\right]$$

$$\left(\frac{(x-\mu)^2}{2\sigma^2}\right) = t$$

$$-\left[\log\frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} p(x)\,dx - \frac{1}{2\sigma^2}\int_{-\infty}^{\infty} (x-\mu)^2 p(x)\,dx\right]$$

$$\log\left(\sqrt{2\pi\sigma^2}\right) + \frac{1}{2}$$

$$\frac{1}{2}\left(\log(2\pi\sigma^2) + 1\right)$$

$$H[x] - H[x/y]$$

$$H_x(x,y) = \int p(x,y) \log(p(x,y))\,dy$$

$$-\underbrace{\iint p(x,y)\log(p(x,y))\,dx\,dy}$$

$$p(x,y) = p(y/x)\,p(x)$$

$$\left[\because\ p(y/x) = \frac{p(x,y)}{p(x)}\right]$$

$$H(x,y) = -\iint p(x,y)\log(p(y/x))\,dx\,dy$$

$$-\iint p(x,y)\log(p(x))\,dx\,dy$$

$$H(x,y) = H(y/x) + H(x)$$

$$\text{|||y} \quad H(x/y) + H(y)$$

$$I(x,y) = KL(p(x,y)\,||\,p(x)p(y))$$

$$= -\iint p(x,y)\log\left(\frac{p(x)(p(y))}{p(x,y)}\right)dx\,dy$$

$$I(x,y) = -\iint p(x,y)\log(p(\cancel{x})\,\cancel{p(y/x)})\,dx\,dy + \iint p(x,y)\log(p(y))\,dx\,dy$$

$$+\iint p(x,y)\log(p(\cancel{x})\,p(y/x)) \Rightarrow H(y/x) + H(y)$$

$\tilde{x} : N(\mu, \sigma^2)$

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

$$\Rightarrow \max_x \left( \frac{(x-\mu)^2}{2\sigma^2} \right) ; \ \mu = x$$

$$f\left(\sum_{i=1}^{M} \lambda_i x_i\right) \leq \sum_{i=1}^{M} \lambda_i f(x_i)$$

$f(x) = \ln x$

$$\Rightarrow \ln\left(\sum \lambda_i x_i\right) \leq \sum \lambda_i \ln(x_i)$$

$\Rightarrow$

$$\sum \lambda_i x_i \leq \prod x_i^{\lambda_i}$$

if $\lambda_i = \frac{1}{n}$

$$\frac{1}{n} \sum x_i \leq \prod x_i^{1/n}$$

AM ≤ GM

7500 $\sigma = 150$ $\mu = 750$

0.99957094

0.9991 42906 6188

99.9%

$$\int_{7000}^{8000} \frac{1}{\sqrt{2\pi}\,150} e^{\frac{(x-\mu)^2}{2\sigma^2}}$$

$$e^{-\frac{1}{2}\|x-x'\|^2}$$

$$\phi_n = \frac{1}{\sqrt{n!}} x^n e^{-x^2/2}$$

Applying inner product

$$e^{-\frac{x'^2}{2}} e^{-\frac{x^2}{2}} \begin{bmatrix} \frac{1\,x}{\sqrt{1!}} \\ \frac{x^2}{\sqrt{2!}} \\ \frac{x^3}{\sqrt{3!}} \\ \vdots \end{bmatrix} \cdot \left[ \frac{1}{\sqrt{1!}} x' \quad \frac{x'^2}{\sqrt{2!}} \quad \frac{x^3}{\sqrt{3!}} \quad \cdots \quad \frac{x^n}{\sqrt{n!}} \right]$$

$$e^{-\left|\frac{x'^2}{2} + \frac{x^2}{2}\right|} \left\{ \frac{1}{1!} x x' + \frac{(xx')^2}{2!} + \cdots \right]$$

$$e^{-\left(\frac{x'^2}{2} + \frac{x^2}{2}\right)} \quad e^{\frac{2xx'}{2}}$$

$$e^{-\frac{1}{2}(x'^2 + x^2 - 2xx')} = e^{-\frac{1}{2}\|x - x'\|^2}$$

$$P(X=x) = a+bx^2 \quad 0 \le x \le 1$$
$$0 \quad \text{otherwise}$$

$$\int_0^1 ax + bx^3 \, dx$$

$$\left. \frac{ax^2}{2} + \frac{bx^4}{4} \right|_0^1$$

$$= \frac{a}{2} + \frac{b}{4} = \frac{3}{5}$$

also $\displaystyle\int_0^1 a + bx^2 \, dx$

$$\left. ax + \frac{bx^3}{3} \right|_0^1 = a + \frac{b}{3} = 1$$

$$\frac{a}{2} + \frac{b}{6} = \frac{1}{2}$$

$$\frac{a}{2} + \frac{b}{4} = \frac{3}{5}$$

$$\frac{6}{10} - \frac{5}{10} = \frac{1}{10} = \frac{3b}{12} - \frac{2b}{12}$$

$$\frac{6}{5}$$

$$\frac{\cancel{6}2}{\cancel{15}5} + a \quad a = \frac{3}{5} \quad b = \frac{6}{5}$$

r g b

3A 3 1ap

40 3 10

3L 4 0L

(0.2) (0.6) (0.2)

$$(0.2)\frac{3}{10} + 0.6\times\frac{3}{10} + 0.\cancel{2}\times\frac{1}{\cancel{2}}$$

$$\frac{6}{100}+\frac{18}{100}+\frac{10}{100} \qquad \frac{34}{100}$$

$$P(G/O) = \frac{P(O/G)\cdot P(G)}{P(O)}$$

$$P(O) = \frac{8}{100}+\frac{18}{100}+\frac{10}{100} = \frac{36}{100}$$

①

$$\frac{\frac{3}{\cancel{10}}\times 0.6}{\frac{\cancel{36}\,12}{\cancel{100}}} \qquad \frac{6}{12} = \frac{1}{2}$$

$x_1, x_2, \dots x_n$

$$P(x;\eta) : \frac{1}{\eta}\left(e^{\frac{-x}{\eta}}\right)$$

$$\max\left(\prod_{i=1}^{n} \frac{1}{\eta} e^{\frac{-x}{\eta}}\right) = \max\left(\frac{1}{\eta^n} e^{\frac{-\sum x_i}{\eta}}\right)$$

$$\Rightarrow \max_{\eta}\left(-n\log\eta - \left(\frac{\sum x_i}{\eta}\right)\right)$$

$$\Rightarrow \min_{\eta}\left(\eta\log\eta + \frac{\sum x_i}{\eta}\right)$$

$$\Rightarrow \log\eta + 1 - \frac{\sum x_i}{\eta^2} = 0$$

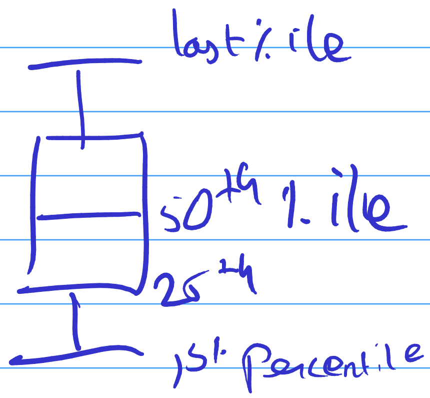
$$\eta^2\log\eta + \eta^2 = \sum x_i$$

$$\eta^2(\log\eta + 1) = \sum x_i$$

$$N\left(5x^2 + 2xy + 3y^2 - 4x + 8y + 4, 1\right)$$

10th order - Lowest } training
1st order - highest }
2nd order - Lowest test error

Since our training points follow a
joint probability distribution
it can be that they are not on the mean
a 10th order polynomial will most likely
overfit in such scenarios




sepal width symmetric

mean is

variance is $\sum (x-\bar{x})^2 p(x)$

variance too, more than mean

Oversampling, dropping columns