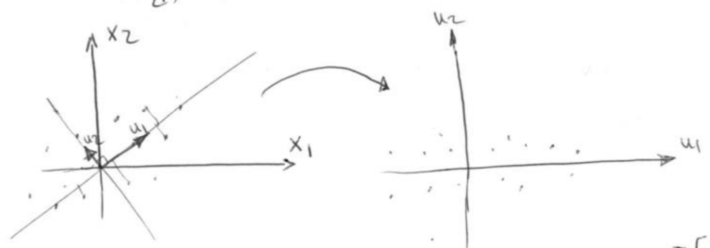
Principle Component Analysis (PCA)

Goal: Reduce dimensionality

Approach: Transform (linearly) the data, into a new coordinate system, where the projection on the first coordinate (PC1), has the highest varience, projection to second coordinate (PC2), has the second highest varience, ... PC_d, has the lowest varience.



Method: For the first coordinate u, given $X = \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^m \end{bmatrix}$ $\rightarrow [x_1^i, x_2^i \cdots x_d^i]$

$$\underset{u}{\text{Arg Max}} \sum_{i=1}^{m} (x^i \cdot u)^2 = (Xu)^T (Xu)$$

$$= u^T X^T X u \quad , \underbrace{\text{s.t } \|u\|_2 = 1}$$

$$= u^T A u$$

the covarience Matrix

ensures a unique answer, where the size of u deosn't play a role in finding Arg Max

Finding $\underset{u}{\text{ArgMax}}\ u^T A u$ is equivalent to

Finding eigenvectors of A:

$$Au = \lambda u \Rightarrow Au - \lambda u = \phi$$

$$(A - \lambda I) u = \phi$$

Steps to find eigen vectors of X

1) center X around zero
2) calculate the covarience ~~of~~ matrix A
3) Find ~~Eigenvectors and~~ eigen values by solving $|A - \lambda I| = \phi$
4) Plug in each eigen value from "3)" into $(A - \lambda I)u = \phi$ to get the corresponding eigen vector.

Example:

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 3 \\ 2 & 2 \\ 1 & 3 \\ 2 & 2 \end{bmatrix}$$

1) center X around zero: $\mu_1 = \frac{0+1+2+1+2}{5} = 1.2$

$$\mu_2 = \frac{1+3+2+3+2}{5} = 2.2$$

$$X - \begin{bmatrix} \mu_1 & \mu_2 \\ \mu_1 & \mu_2 \\ \mu_1 & \mu_2 \\ \mu_1 & \mu_2 \\ \mu_1 & \mu_2 \end{bmatrix} = \begin{bmatrix} -1.2 & -1.2 \\ -0.2 & 0.8 \\ 0.8 & -0.2 \\ -0.2 & 0.8 \\ 0.8 & -0.2 \end{bmatrix} \text{: new } X$$

2) calculate A (the covarience matrix) =

$$\mathrm{Cov}(X) = \frac{X^T X}{N-1} = \cdots = \begin{bmatrix} 0.7 & 0.2 \\ 0.2 & 0.7 \end{bmatrix} = A$$

scaling factor (not necessary)

3) Find eigen values: $|A - \lambda I| = \phi$

$$\begin{vmatrix} 0.7-\lambda & 0.2 \\ 0.2 & 0.7-\lambda \end{vmatrix} = (0.7-\lambda)^2 - (0.2 \times 0.2) = \phi$$

$$= \lambda^2 - 1.4\lambda + 0.45 = \phi$$

$$\Rightarrow \begin{cases} \lambda_1 = \underline{0.9} \text{ (largest eigen value)} \\ \lambda_2 = 0.5 \end{cases}$$

4) Find eigen vectors.

- $\lambda_1 = 0.9$

$$(A - \lambda_1 I) u = \begin{bmatrix} 0.7-0.9 & 0.2 \\ 0.2 & 0.7-0.9 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \phi$$

$$\Rightarrow \begin{cases} -0.2u_1 + 0.2u_2 = \phi \\ 0.2u_1 - 0.2u_2 = \phi \end{cases}$$

$$\Rightarrow u_1 = u_2 \Rightarrow u = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \dots$$

but $\underset{\text{constraint}}{\|u\| = 1} \Rightarrow u_{\lambda_1} = \begin{bmatrix} 0.7 \\ 0.7 \end{bmatrix}$

first eigen vector

- $\lambda_2 = 0.5$

can find <u>second</u> eigen vector similarly

hence $PC1 = X u_{\lambda_1}$ | $u_{\lambda_1} \perp u_{\lambda_2}$

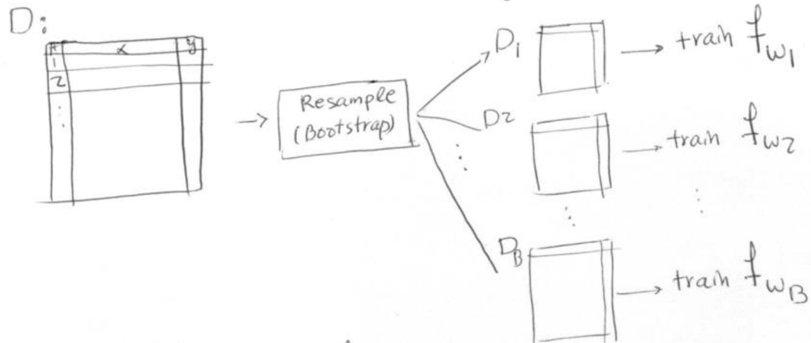
$PC2 = X u_{\lambda_2}$ | $Var(PC1) \geq Var(PC2)$

Bootstrap Method for providing Confidence Intervals

Goal: Instead of point prediction, provide an interval to achieve the requested confidence (e.g. 95%)

Method: Assume $f_w(x) = \hat{y}$, is a regressor with parameter w, which gives $\hat{y}$ prediction given x.

1) Training:



2) Providing confidence interval:

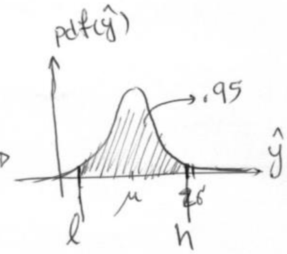
given x: $\hat{y}_1 = f_{w_1}(x)$

$\hat{y}_2 = f_{w_2}(x)$

$\vdots$

$\hat{y}_B = f_{w_2}(x)$

$$\left.\begin{aligned} \mu &= \text{mean}(\hat{y}_1, \hat{y}_2, \cdots \hat{y}_B) \\ \sigma^2 &= \text{variance}(\hat{y}_1, \hat{y}_2, \cdots \hat{y}_B) \end{aligned}\right\} \rightarrow$$



predict $l \leq \hat{y} \leq h$ as confidence interval