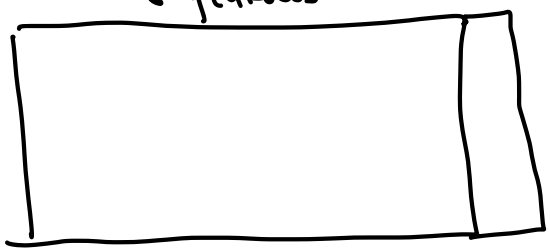
5 quizzes 2 midterms final 8 features

8 features

20 students

20 students



Scores

$\sum .1 q_i + \sum .2 m + .1 f$

score

Scores

- A (linear) "score" is an artificially created feature that is a linear combination of existing features.

$$S_j = \sum a_i x_{ji}$$

- A score is defined by k weights

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix}$$

and the values of a score on the samples is computed by

$$\begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_n \end{bmatrix} = X_0 \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix}$$

$$\underline{\underline{S_i}} = \sum_j x_{ij} a_j$$

$$\sum_i S_i = \sum_i \sum_j x_{ij} a_j$$

$$= \sum_j \left(\sum_i x_{ij} \right) a_j \quad = 0 \text{ since } \sum_{i=1}^{N} x_{ij} = 0$$

Mean of scores

A linear combination of features with mean zero has mean zero.

Variance and Covariance of scores

If S is a score corresponding to a weight vector a, then

$$\sigma_S^2 = a^\intercal D_0 a.$$

If S and T are two scores corresponding to weights a and b, then the covariance of S and T is given by

$$\sigma_{ST} = b^\intercal D_0 a.$$

$$\underline{S = X_0 a} \qquad a = \begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix}$$

S is an $N \times 1$ vector entries are the values of the score for each row of X_0

$$\begin{bmatrix} s_1 \\ \vdots \\ s_N \end{bmatrix} \qquad S \cdot S = \sum s_i^2$$

$$\underbrace{S^T S}$$

$$\sigma_S^2 = \frac{1}{N} S^T S$$

$$S = X_0 a$$

$$S^T = (X_0 a)^T = a^T X_0^T$$

$$\sigma_S^2 = \frac{1}{N} a^T X_0^T X_0 a$$

$$= a^T \underline{\left(\frac{1}{N} X_0^T X_0\right)} a$$

$$= a^T D_0 a$$

$$a = \begin{bmatrix} a_1 \\ \vdots \\ a_N \end{bmatrix} \qquad b = \begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix}$$

$$S = X_0 a \qquad T = X_0 b$$

$$\sigma_{ST} = \frac{1}{N} \sum s_i t_i = \frac{1}{N} S \cdot T = \frac{1}{N} S^T T = \frac{1}{N} a^T X_0 X_0 b$$

$$= a^T D_0 b$$