

## Supervised pca

Saturday, May 26, 2018 10:28 AM

There are 2 versions for supervised PCA

### Version 1

Given data  $X \in \mathbb{R}^{n \times d}$  and its corresponding labels  $Y \in \mathbb{R}^{n \times 1}$ . Assuming that  $X$  is centered.  $n$  is # of samples,  $d$  is number of features.

Step 1. Normalize each column of  $X$  into  $\hat{X}$

Step 2. Project  $Y$  onto  $\hat{X}$ , this gives us the relationship between each feature to  $Y$ . We get rid of the lower relationship features, so that  $X$  only have  $q$  columns.

Note: This implies that we remove features that are not related to  $Y$

Step 3. Using reduced dimensioned  $X$ , perform PCA on it.

### Version 2

Given data  $X \in \mathbb{R}^{n \times d}$  and  $Y$  as the label.

We want to find a projection that maximize  $XW$  and  $Y$ . We use HSIC.

$$\max_W \text{Tr}(K_X W^T H K_Y H)$$

If we use a linear kernel for  $XW$

$$\max_W \text{Tr}(X W W^T X^T H K_Y H)$$

We can solve  $W$  with

$$\max_W \text{Tr}(W^T X^T H K_Y H X W)$$

$$\text{s.t. } W^T W = I$$