Figure 1: Adding new features for linear classification

Content 7
Building new features to be used by linear classifiers K Nearest Neighbor Algorithm Using a validation data set to address the overfitting issue

1 Keywords

- A linear classifier means that the predicted output is a linear combination of base functions.

- Example of base functions being used in a linear classifier.

$$\widehat{y} = \mathbf{1} \left(\sum_{\substack{\mathcal{F}' \subset \mathcal{F} \\ \#\mathcal{F}' \leq d}} a_{\mathcal{F}'} \prod_{f \in \mathcal{F}'} x_f + b > 0 \right) \tag{1}$$

where b is the intercept, $a_{\mathcal{F}'}$ are the weights and $\#\mathcal{F}'$ is the size of set $\mathcal{F}'$.

- Euclidean distance

$$d(\mathbf{x}, \mathbf{x}') = \sqrt{\sum_f (x_f - x'_f)^2} \tag{2}$$

- kNN

- RBF kernel

$$x_{n,f} = e^{-\gamma(\sum_{f'}(x_{n,f'} - \tilde{x}_{f,f'})^2)} \tag{3}$$

where $\tilde{\mathbf{x}}_f$ are the selected samples.

- Validation dataset.

1