Rafael Antonio Garcia Mar

Rafael Garcia

Feb 21, 2020

Part A: Theory

Linear Regression and Regularization

The loss function that includes the L_2 penalty is given by

$$\mathcal{L} = \sum_i (\vec{w}^T \vec{x}_i - y_i)^2 + \lambda |\vec{w}|_2^2$$

In terms of X (the matrix whose rows are $\{\vec{x}_i\}$) and $\vec{y}$:

$$\mathcal{L} = (X\vec{w} - \vec{y})^T (X\vec{w} - \vec{y}) + \lambda \vec{w}^T \vec{w} = \vec{w}^T X^T X \vec{w} - \vec{w}^T X^T \vec{y} - \vec{y}^T X \vec{w} + \vec{y}^T \vec{y} + \lambda \vec{w}^T \vec{w}$$

Minimizing $\mathcal{L}$:

$$\frac{\partial \mathcal{L}}{\partial \vec{w}^T} = 2X^T X \vec{w} - 2X^T \vec{y} + 2\lambda \vec{w} = 0$$

$$(X^T X + \lambda \mathbb{1})\vec{w} = X^T \vec{y}$$

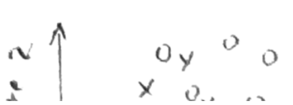
$$\vec{w} = (X^T X + \lambda \mathbb{1})^{-1} X^T \vec{y}$$

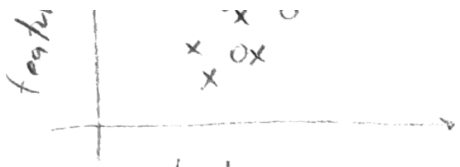
Solution with no regularization: $\vec{w} = (X^T X)^{-1} X^T \vec{y}$

The solution is changed by the term $\lambda \mathbb{1}$. This penalty prevents the values of $\vec{w}$ from being too high. Higher values of $\vec{w}$ are usually due to overfitting. The penalty then prevents overfitting.

Density Estimation

n = number of data points

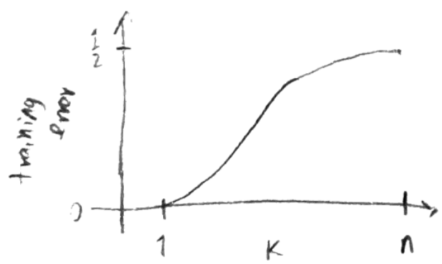




KNN with $k = n$: A point has $\frac{n}{2}$ neighbors from each class. Then the training error is $\frac{1}{2}$.

KNN with $k = 1$: A point from the training set will be its own neighbor. Then the training error is 0.

The training error would look like



The monotonic behaviour is not guaranteed.

Feature Selection and Preprocessing

There are two main things that could be improved:

1) They are selecting the features set $\mathcal{F}$ by minimizing the training error

using all of the data D. They should use a train-test split and cross-validation to minimize the validation error and safeguard against random chance.

2) The validation procedure is done on a subset $D' \subset D$. The set F was selected by using all of D. There is a train-test contamination. The feature set F will perform good on D' because it was selected based on $D \subset D'$. They sould perform this validation on new data.