1 Regression Loss Functions

1.1 L^p Norm

Let $\boldsymbol{y}(\boldsymbol{w}, \boldsymbol{b}) : \mathbb{R}^n \to \mathbb{R}^m$ be a model defined by $\boldsymbol{w}$ and $\boldsymbol{b}$ that map $\boldsymbol{x} \in \mathbb{R}^n$ into $\boldsymbol{y} \in \mathbb{R}^m$. Mathematically the L^1 norm is

$$\|\boldsymbol{y}(\boldsymbol{w}, \boldsymbol{b}) - \hat{\boldsymbol{y}}\|_1 = \sum_{k=1}^{m} \left| y_k(\boldsymbol{w}, \boldsymbol{b}) - \hat{y}_k \right|$$

The L^p norm is

$$\|\boldsymbol{y}(\boldsymbol{w}, \boldsymbol{d}) - \hat{\boldsymbol{y}}\|_p = \left(\sum_{k=1}^{m} \left(y_k(\boldsymbol{w}, \boldsymbol{b}) - \hat{y}_k \right)^p \right)^{1/p}$$

The L^∞ norm is

$$\|\boldsymbol{y}(\boldsymbol{w}, \boldsymbol{b}) - \hat{\boldsymbol{y}}\|_\infty = \max_k |y_k(\boldsymbol{w}, \boldsymbol{b}) - \hat{y}_k|$$

1.2 Mean Absolute Error(L^1 Loss)

Suppose there are N instances $\{\boldsymbol{x}_i\}_{i=1}^N$. The MAE is defined by

$$L(\boldsymbol{w}, \boldsymbol{b}) = \frac{\sum_{n=1}^{N} \|\boldsymbol{y}^n(\boldsymbol{w}, \boldsymbol{b}) - \hat{\boldsymbol{y}^n}\|_1}{N}$$

1.3 Mean Square Error(L^2 Loss)

Suppose there are N instances $\{\boldsymbol{x}_i\}_{i=1}^N$. The MSE is defined by

$$L(\boldsymbol{w}, \boldsymbol{b}) = \frac{\sum_{n=1}^{N} \|\boldsymbol{y}^n(\boldsymbol{w}, \boldsymbol{b}) - \hat{\boldsymbol{y}^n}\|_2^2}{N}$$

1.4 Regularization

Add new term

$$\tilde{L}(\boldsymbol{w}, \boldsymbol{b}) = L(\boldsymbol{w}, \boldsymbol{b}) + \lambda \|\boldsymbol{w}\|_2^2$$

The last term will make loss function smoother since we also minimize weights.