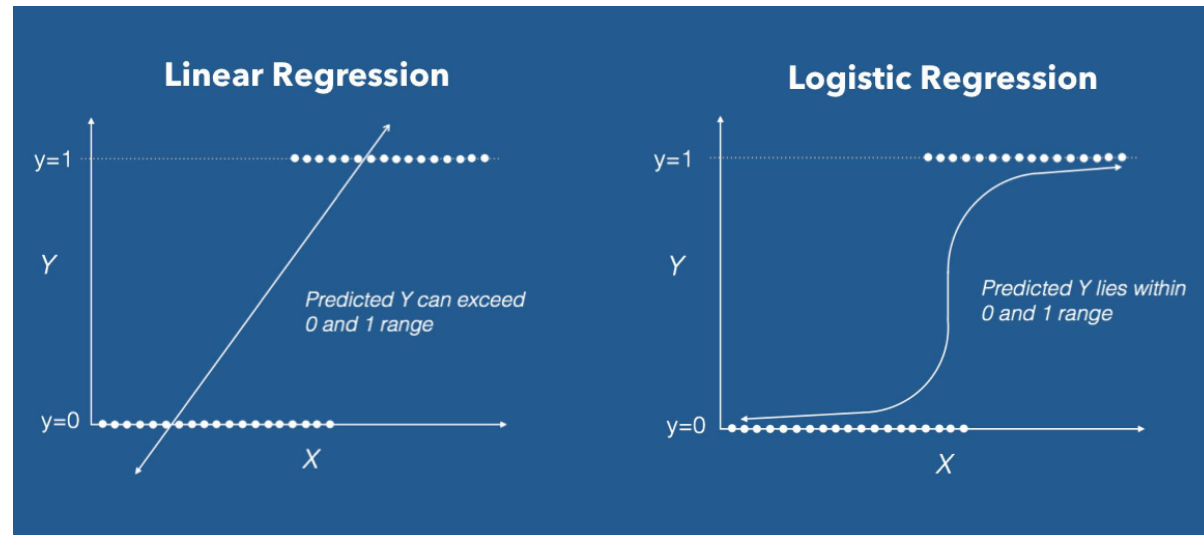


LESSON 3: REGULARIZATIONS



1. Regularization introduction

One simple way to reduce overfitting model is adding a term into loss function

$$J_{\text{reg}}(w) = J(w) + \lambda R(w)$$

the $J_{\text{reg}}(w)$ is called regularized loss function.

Minimizing the $J_{\text{reg}}(w)$ is minimizing both $J(w)$ term and $R(w)$ term.

While minimizing $J(w)$ is the main objective, minimizing $R(w)$ will make some weights in our model become near to zero or, in other words, deactivate these weights.

Deactivating some weights will degrade the complexity of our model and it help us reduce overfitting.

Sometimes, we call "deactivating some weights" as **weights decay**.

2. Popular regularization function

2.1. Overall formula of Norm

norm p of x is calculated by

$$||x||_p = (|x_1|^p + |x_2|^p + \dots + |x_n|^p)^{\frac{1}{p}}$$

2.2. L1 regularization

$$R(w) = ||w||_1 = |w_1| + |w_2| + \dots + |w_n|$$

$||w||_1$ is called **norm 1**

Lasso regression

In Linear regression, we can add L1 regularization into the loss function to create **Lasso regression**.

Original MSE for Linear regression,

$$\begin{aligned}
 &MSE(w) \\
 &= \frac{1}{N} \sum_{i=1}^N (\hat{y}^{(i)} - y^{(i)})^2 \\
 &= \frac{1}{N} ||\hat{y} - y||_2^2
 \end{aligned}$$

MSE for Lasso regression,

$$MSE(w) = \frac{1}{N} ||\hat{y} - y||_2^2 + ||w||_1$$

2.3. L2 regularization

$||w||_2$ is called **norm 2**

Ridge regression

In Linear regression, we can add L2 regularization into the loss function to create ***Ridge regression***.

MSE for Ridge regression,

$$MSE(w) = \frac{1}{N} ||\hat{y} - y||_2^2 + ||w||_2^2$$

In []: