Machine Learning Diploma

Level3: Machine Learning

Session 3

AMIT

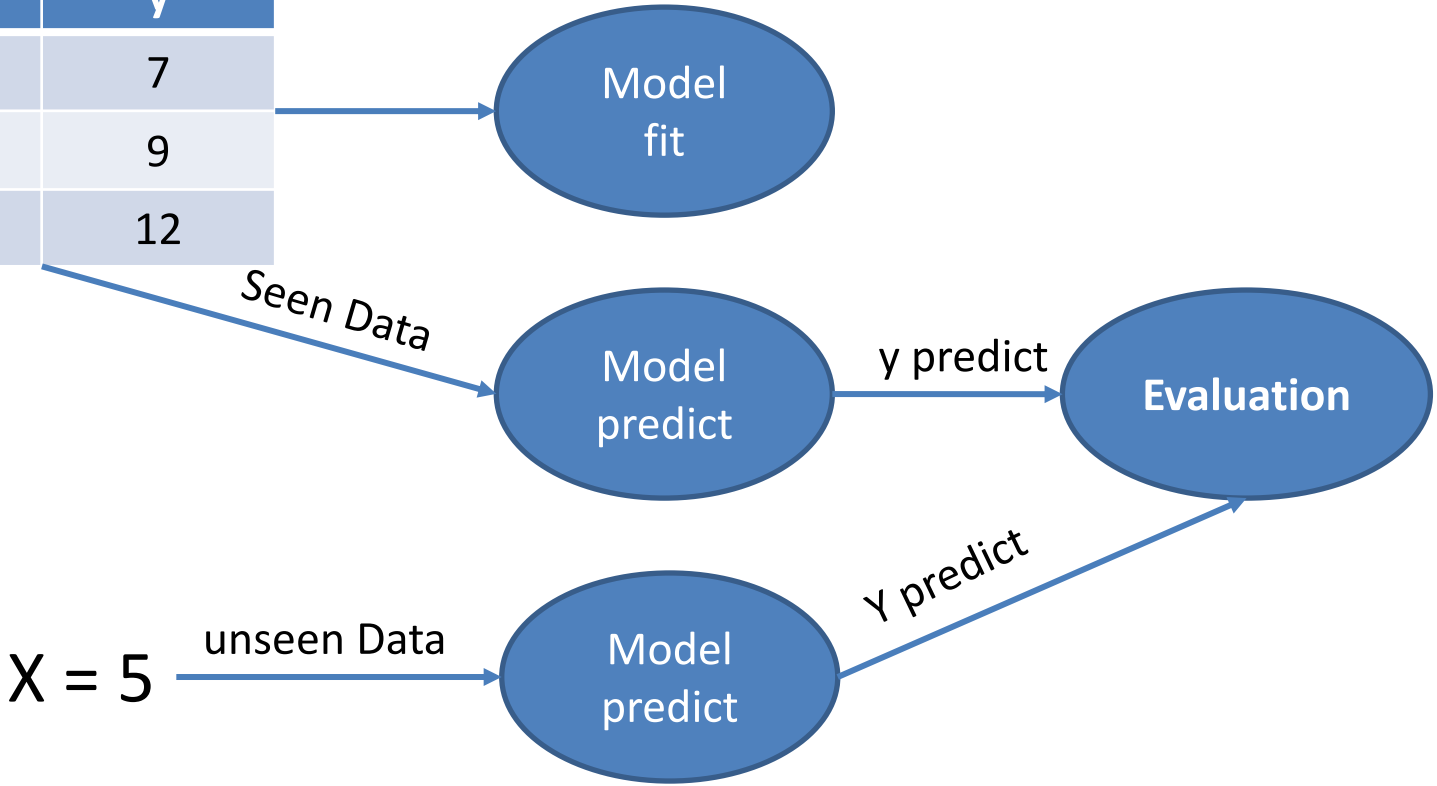
Agenda

- ➔ Training, Validation, and Testing Datasets
- ➔ Error In ML models
- ➔ Bias and Variance
- ➔ Overfitting and Underfitting
- ➔ Regularized Linear Regression

AMIT

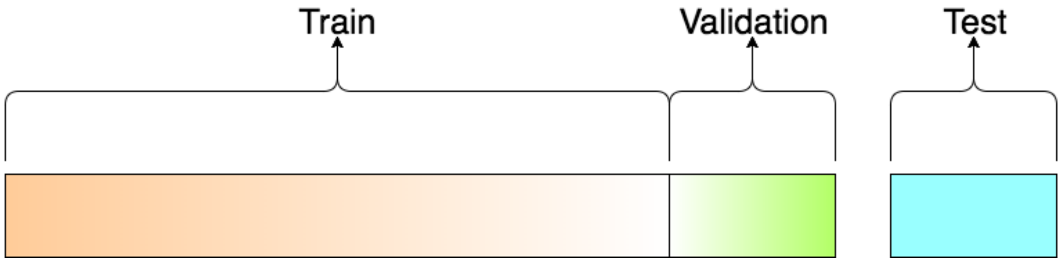
1. Training, Validation, and Testing Datasets

x	y
3	7
4	9
6	12


Model fit
Seen Data
Model predict
y predict
Evaluation
X = 5
unseen Data
Model predict
Y predict

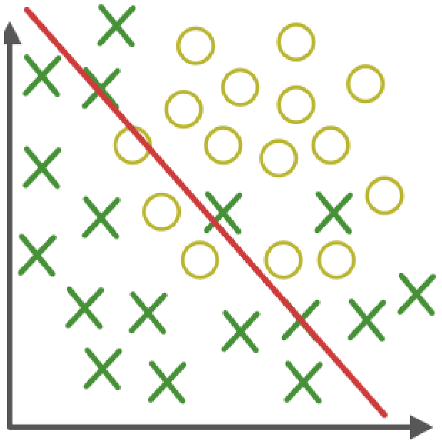
Definition:

- ➔ **Training Dataset**: The sample of data used to fit the model.
- ➔ **Validation Dataset**: The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyper parameters.
- ➔ **Testing Dataset**: The sample of data used to provide an unbiased evaluation of a final model fit on the training dataset.



2. Error in Ml models

Error in ML Models

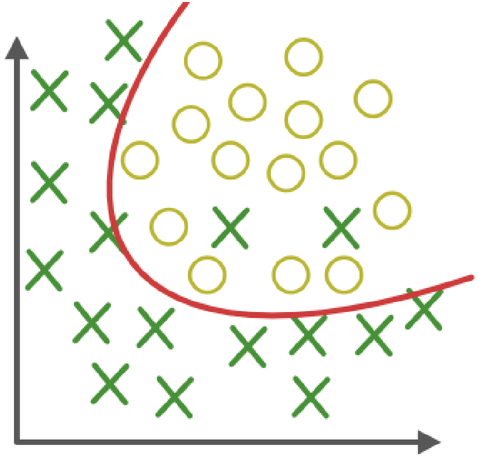


1

Error 7/32

True 25/32

Accuracy = 78 %

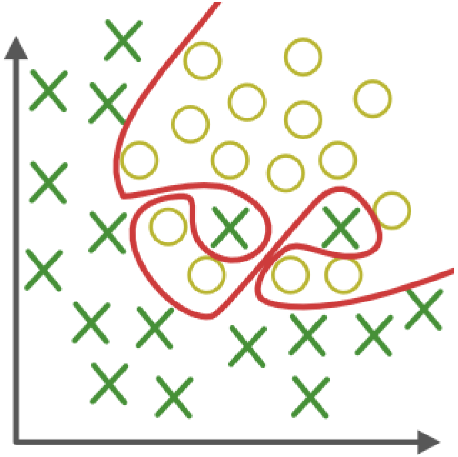


2

Error 2/32

True 30/32

Accuracy = 94 %



3

Error 0/32

True 32/32

Accuracy = 100 %

Error in ML Models

Y and other covariates as X. We assume there is a relationship between the two such that Where e is the error term.

$$Y = f(X) + e$$

We will make a model f^(X) of f(X) using linear regression or any other modeling technique.

$$Err(x) = E\left[(Y - \hat{f}(x))^2\right]$$

The Err(x) can be further decomposed as

$$Err(x) = \left(E[\hat{f}(x)] - f(x)\right)^2 + E\left[\left(\hat{f}(x) - E[\hat{f}(x)]\right)^2\right] + \sigma_e^2$$

$$Err(x) = \mathrm{Bias}^2 + \mathrm{Variance} + \mathrm{Irreducible\ Error}$$

AMIT

3. Bias and Variance

Bias and Variance

Bias is the difference between the average prediction of our model and the correct value which we are trying to predict.
Model with high bias pays very little attention to the training data and oversimplifies the model. It always leads to high error on training and test data.

Characteristics of a high bias model include:

•Failure to capture proper data trends
•Potential towards underfitting
•More generalized/overly simplified
•High error rate

AMIT

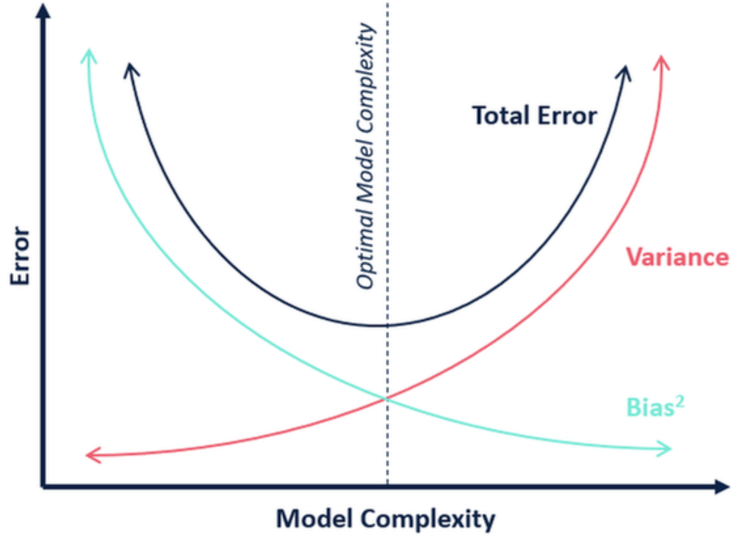
Bias and Variance

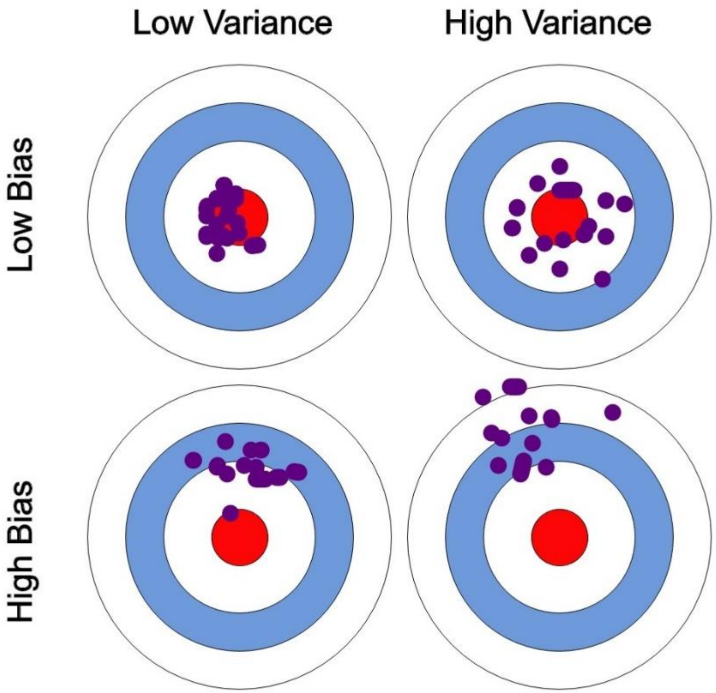
Variance is the variability of model prediction for a given data point or a value which tells us spread of our data.
Model with high variance pays a lot of attention to training data and does not generalize on the data which it hasn't seen before.

Characteristics of a high variance model include:

- Noise in the data set
- Potential towards overfitting
- Complex models
- Trying to put all data points as close as possible

Bias and Variance





Bias and Variance

Ways to reduce High Bias:

- Increase the input features as the model is underfitted.
- Decrease the regularization term.
- Use more complex models, such as including some polynomial features.

Ways to reduce high variance:

- Reduce the input features or number of parameters as a model is overfitted.
- Do not use a much complex model.
- Increase the training data.
- Increase the Regularization term.

4. Overfitting and Underfitting

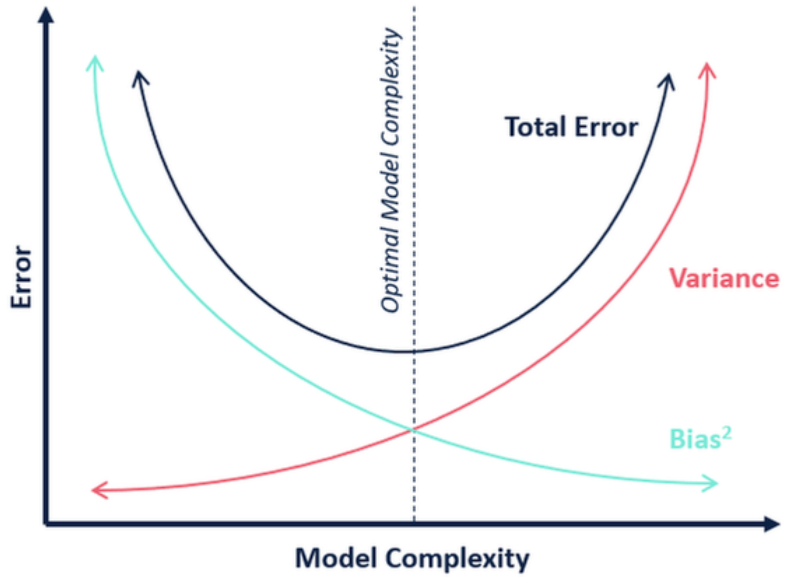
Definition:

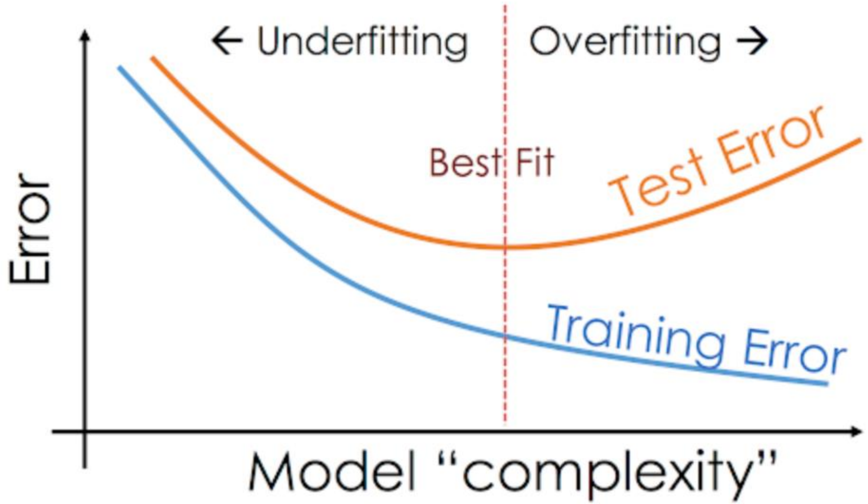
- ➔ **Overfitting:** Good performance on the training data, poor generalization to other data.
- ➔ **Overfitting:** Happens when a model learns the detail and noise in the training data to the extent that it negatively impacts the performance of the model on new data.
- ➔ **Underfitting:** Poor performance on the training data and poor generalization to other data.

Dataset	Underfitting	Optimum fitting	Overfitting
Train	Low	High	High
Test	Low	High	Low

AMIT

Overfitting and Underfitting:





Overfitting and Underfitting:

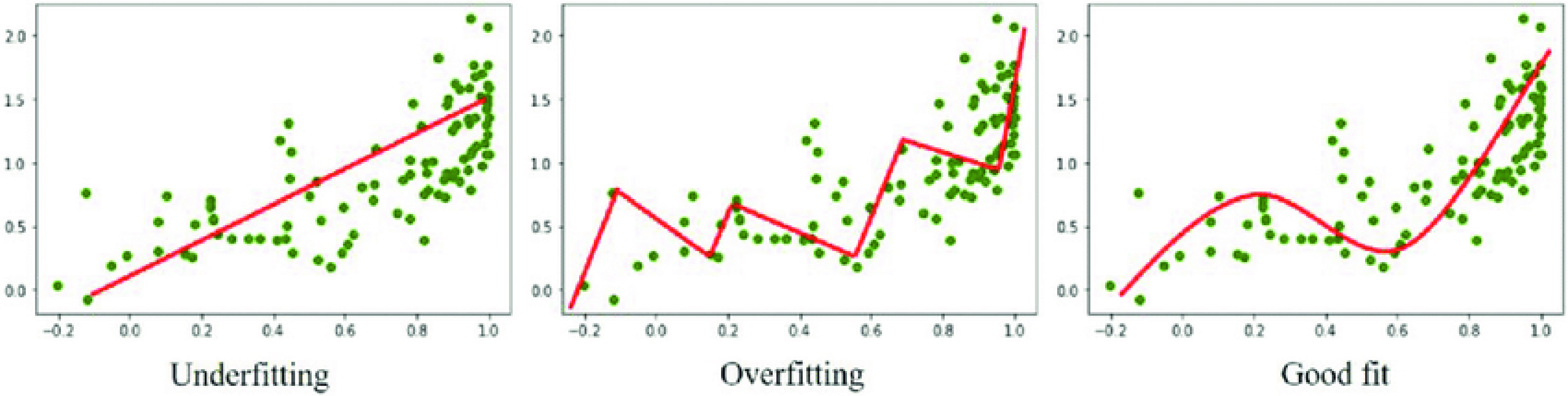
Techniques to reduce underfitting:

1. Increase model complexity
2. Increase the number of features, performing feature engineering
3. Remove noise from the data.
4. Increase the number of epochs or increase the duration of training to get better results.

Techniques to reduce overfitting:

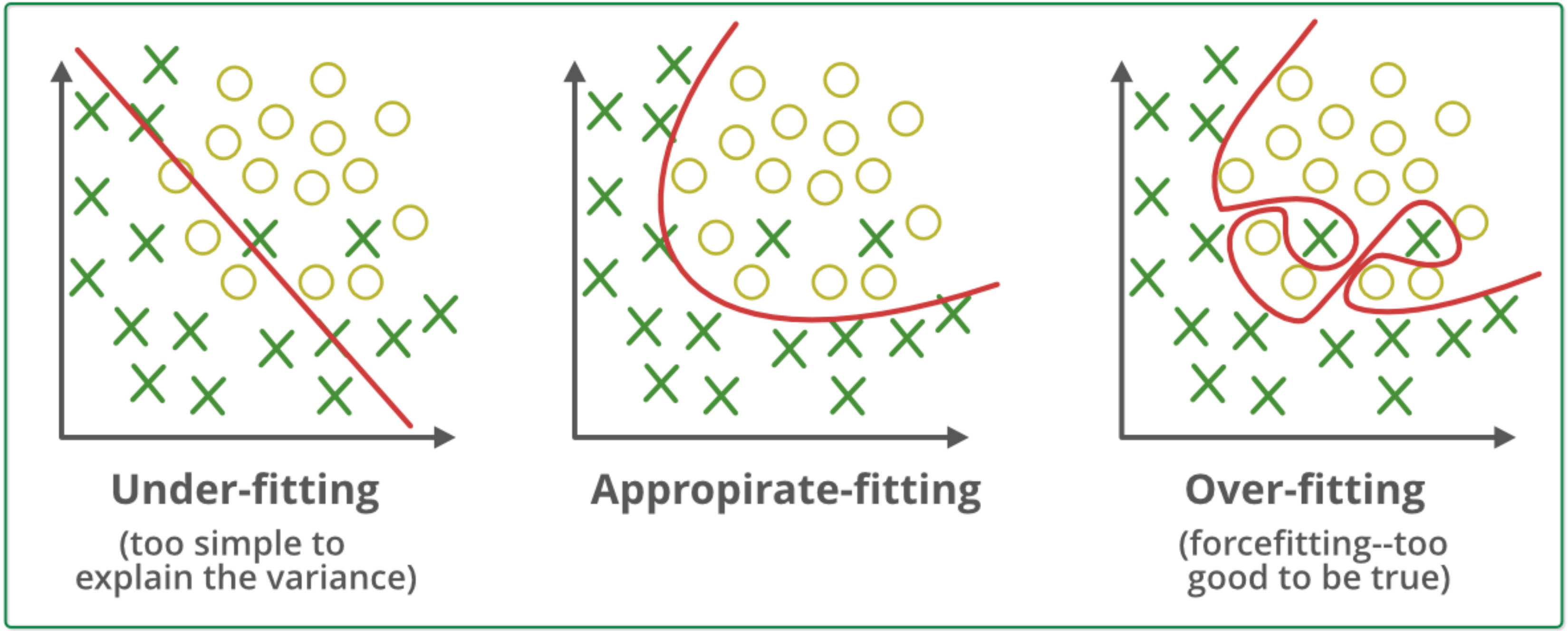
1. Increase training data.
2. Reduce model complexity.
3. Early stopping during the training phase (have an eye over the loss over the training period as soon as loss begins to increase stop training).
4. Ridge Regularization and Lasso Regularization
5. Use dropout for neural networks to tackle overfitting

Example:



2.0
1.5
1.0
0.5
0.0
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
Underfitting
Overfitting
Good fit

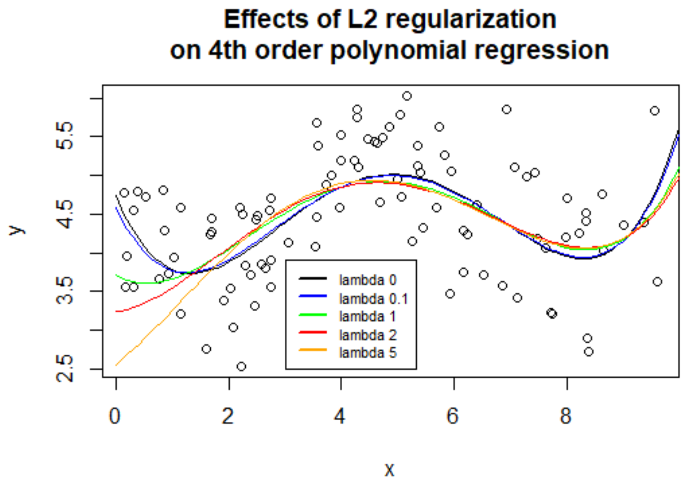
Overfitting and Underfitting:



5. Regularized Linear Regression

Regularization:

- ➔ **Regularization** is a technique to reduce overfitting.
- ➔ **Regularized Regression**: This is a form of regression, that constrains or regularizes the coefficient estimates towards zero. In other words, this technique discourages learning a more complex or flexible model, so as to avoid the risk of overfitting.



Regularization Parameter:

- **Regularization** is a technique used for tuning the function by adding an additional penalty term in the error function. The additional term controls the excessively fluctuating function such that the coefficients don't take extreme values.
- **The Regularization Parameter** is a control on your fitting parameters. As the magnitudes of the fitting parameters increase, there will be an increasing penalty on the cost function.
- Increasing **lambda** results in less overfitting but also greater bias.

$$\text{minimize}(\text{Loss}(\text{Data}|\text{Model}) + \lambda\,\text{complexity}(\text{Model}))$$

AMIT

Types of Regularization:

- ➔ **L2** and **L1** are the most common types of regularization. Regularization works on the premise that smaller weights lead to simpler models which in results helps in avoiding overfitting. So to obtain a smaller weight matrix, these techniques add a 'regularization term' along with the loss to obtain the cost function.
- ➔ **L1 Regularization** gives output in binary weights from 0 to 1 for the model's features and is adopted for decreasing the number of features in a huge dimensional dataset.
- ➔ **L2 Regularization** disperse the error terms in all the weights that leads to more accurate customized final models.

AMIT

Types of Regularization:

L1 Regularization

$$\text{Cost} = \sum_{i=0}^{N}\left(y_i - \sum_{j=0}^{M} x_{ij}W_j\right)^2 + \lambda \sum_{j=0}^{M} |W_j|$$

L2 Regularization

$$\text{Cost} = \underbrace{\sum_{i=0}^{N}\left(y_i - \sum_{j=0}^{M} x_{ij}W_j\right)^2}_{\text{Loss function}} + \underbrace{\lambda \sum_{j=0}^{M} W_j^2}_{\text{Regularization Term}}$$

Any Questions?

THANK YOU!