CSE 471: Machine Learning

Machine Learning

Tajkia Nuri Ananna

February 22, 2023

Lecturer, Metropolitan University

Table of contents

Gradient Descent

Cost Function

Cost Function

- Suppose, Y is sales, X is advertising spend and we want to estimate how advertising spend impacts sales
- Definition: Measure of how wrong the model is in terms of its ability to estimate the relationship between X and y
- Cost function is expressed as difference or distance between the predicted value and the actual value.
- Target of Gradient Descent: minimizing the cost function

Example of Cost Function

Example of Cost Function

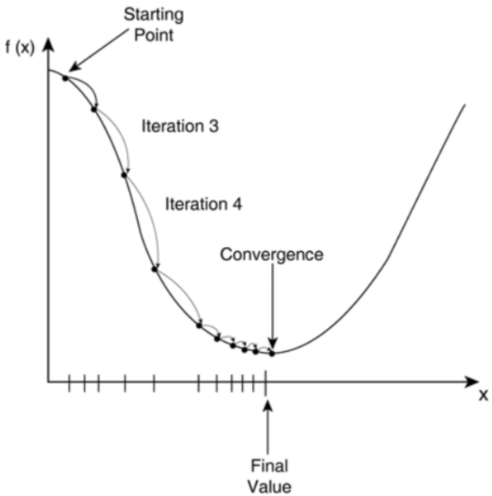
- As children, we learn to identify right and good by being told not to do certain things or being punished for having done something we shouldn't
- Imagine a four year old sitting next to fire, but without knowing the danger of fire she puts her finger into it and gets burned
- The next time she sits by the fire, she doesn't get burned, but she sits too close, gets too hot and has to move away.
- The third time she sits by the fire she finds the distance that keeps her warm without exposing her to any danger.
- Through experience and feedback the kid learns the optimal distance to sit from the fire.
- The heat from the fire in this example acts as a **cost function** — it helps the learner to correct / change behaviour to minimize mistakes.

Gradient Descent

Gradient Descent

- Gradient descent: An efficient optimization algorithm that attempts to find a local or global minima of a function.
- Enables a model to learn the gradient or direction that the model should take in order to reduce errors
- Gradient = Slope (Draw a line at any point of a function, the tangent of the angle between the line and X axis) , Descent = Decrease
- A ML model always wants low error with maximum accuracy, in order to decrease error we will intuit our algorithm that you're doing something wrong that is needed to be rectified, that would be done through Gradient Descent.

Gradient Descent



Gradient Descent

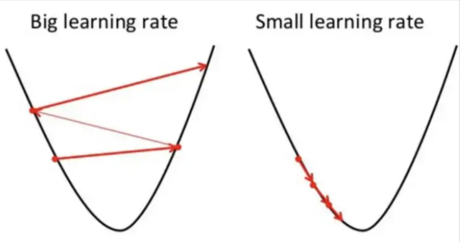
Gradient Descent

- GD estimates the weights of the model in many iterations by minimizing a cost function at every step.
- The Algorithm is
  $w_j = w_j - \alpha \frac{\partial}{\partial w_j} J\left(w_0, w_1\right)$ where α is the learning rate/step size and $J\left(w_0, w_1\right)$ is the cost function

Learning Rate

Learning Rate

- Determines how fast or slow we will move towards the optimal position.
- If the learning rate is very large we will skip the optimal solution.
- If it is too small we will need too many iterations to converge to the best values.



Bias/Variance Trade-off

Overfitting

Overfitting

- Overfitting is a situation when your model is too complex for your data.
- More formally, your hypothesis about data distribution is wrong and too complex — for example, your data is linear and your model is high-degree polynomial.
- Happens when our model captures the noise along with the underlying pattern in data and the model tries to fit exactly to the training data
- Overfitting also happens when the number of features (columns)are more than the number of training data (row)
- As a result, it performs very well with training dataset, but fails to generalize the prediction in case of testing dataset.

Over-fitting

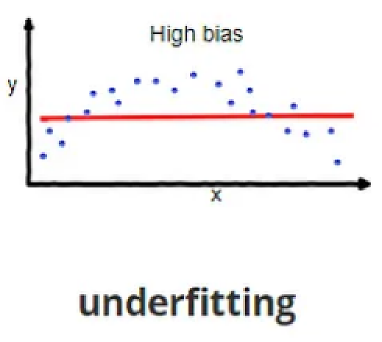
Overfitting

Underfitting

Underfitting

- Underfitting is a situation when your model is too simple for your data.
- More formally, your hypothesis about data distribution is wrong and too simple — for example, your data is quadratic and your model is linear.
- Underfitting happens when a model unable to capture the underlying pattern of the data.
- As a result, it does not perform well even on training data.

Underfitting



Variance

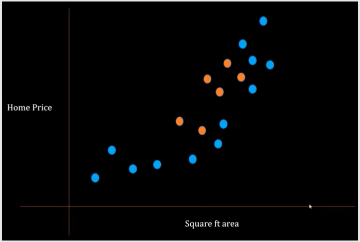
Variance

- 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.
- As a result, such models perform very well on training data but has high error rates on test data.
- Overfitted model has high variance

Overfitting and High Variance

Overfitting and High Variance

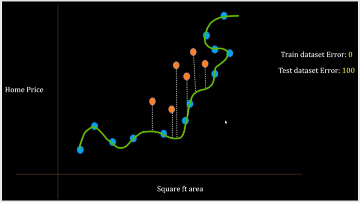
- Suppose, in the following figure blue dots are our training data and orange dots are test data



Overfitting and High Variance

Overfitting and High Variance

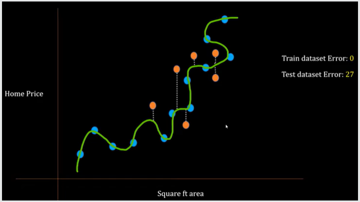
- Here we train your data using a overfitted model and calculate the average test data error
- Suppose, your test error is 100



Overfitting and High Variance

Overfitting and High Variance

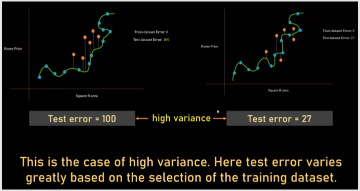
- Suppose, another scientist chosed a different set of training dataset and got a test error of just 27



Overfitting and High Variance

Overfitting and High Variance

- That's why an overfitted model is called high variance



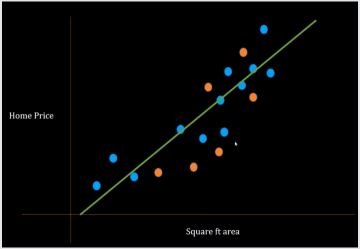
Bias

Bias

- Bias is a measurement of how accurately a model can capture a pattern in a training dataset
- 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.
- Train error is called high bias, because it doesn't matter whatever you do, the model doesn't change it's shape (it is biased)
- Underfitted model has high bias

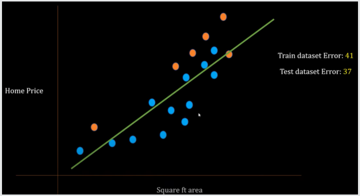
Underfitting and High Bias

- Suppose, you chosed a underfitted model, where the line does not pass all the blue dots, training error is high



Underfitting and High Bias

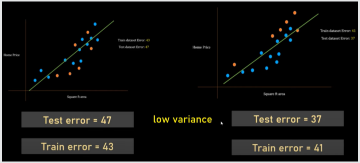
- When you select a different set of training data points, both the cases the line is different but training and test error is similar



Underfitting and Low Variance

Underfitting and Low Variance

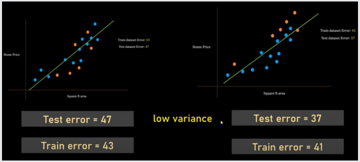
- It is called low variance because, test error does not vary that much



Underfitting and High Bias

Underfitting and High Bias

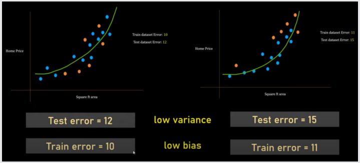
- Bias is a measurement of how accurately a model can capture a pattern in a training dataset. In this case, the training error is big, so High Bias
- Bias means train error, Variance meaning test error
- The higher the train error, the higher the bias



Balanced Fit

Balanced Fit

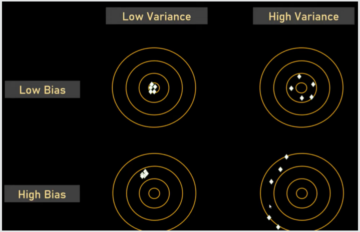
- Test Error does not vary = Low Variance
- Train error is low, Low Bias



Bias Variance Tradeoff (Bulls Eye Diagram

Bias Variance Tradeoff

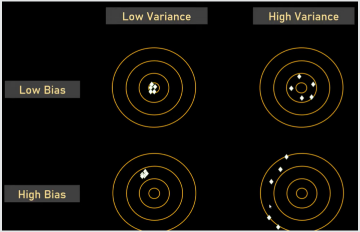
- The center point is the truth
- High variance Low Bias: predicted values are scattered apart but closer to truth
- High Bias High Variance: Far away from truth
- Low Bias: Predicted values near to the truth
- High Bias: Far away from middle circle
- Predicted values clustered together Low variance



Bias Variance Tradeoff

Bias Variance Tradeoff

- Low Variance, High Bias = Underfitting
- Low Bias, High Variance = Overfitting



Bias Variance Tradeoff

Bias Variance Tradeoff

- If a model has low variance, means the model is less complex
- Similarly, Low Bias means highly complex model
- if we want to achieve both together, we have to build a model which is complex and simple at the same time
- This is known as bias variance tradeoff

How to build a balanced fit model

How to build a balanced fit model

- Cross Validation
- Regularization
- Dimensionality Reduction
- Ensemble techniques
- Bagging, Boosting etc


Thank
you!