Deep Learning and Computer Vision

Mentor: Kenny Zhang
Mentee: Angela Zhao

03/18/2021

Topics We’ve Covered

- Brief Review of Linear Regression
- Logistic Regression
- Neural Network
- Convolutional Neural Network
- Bias & Variance Tradeoff
- Python Exercise: Image Classification

How does the computer predict

whether an email is spam or not?

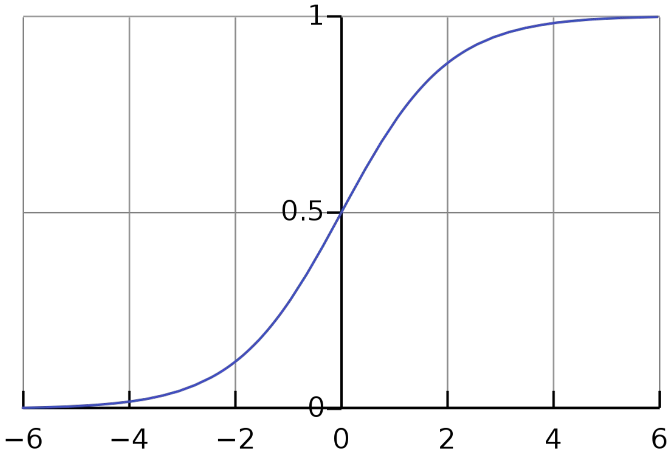
How does the machine conclude

whether it is a benign or malignant tumor?

Logistic Regression

Logistic/Sigmoid equation: $g(x) = \dfrac{1}{1 + e^{-x}}$

Output: (0, 1)



Hypothesis Analysis

Analysis Criteria: How confident that the predicted value is equal to actual value given an input X.

Hypothesis Analysis

Analysis Criteria: How confident that the predicted value is equal to actual value given an input X.

Decision Boundary (Threshold)

- Often set to be 0.5: predicted_value < 0.5, classify into 0; otherwise 1
- For logistic regression, it is linear; for some other ones, it can be nonlinear

Recall: Linear Regression

Hypothesis function: y = ax + b, where X is the input training data and Y is the data label

Cost function: $$J = \frac{1}{n}\sum_{i=1}^{n}(pred_i - y_i)^2$$ → Mean Squared Error

Note: For some statistical models satisfied a certain property, we want to minimize the cost function by taking the derivative with respect to y and then setting it equal to 0 to find where the y-value is.

Apply a similar format aforementioned into the cost function of logistic regression?

$$J = \frac{1}{n}\sum_{i=1}^{n}(pred_i - y_i)^2$$

$$pred_i = \frac{1}{1 + e^{-x_i}}$$

Apply a similar format aforementioned into the cost function of logistic regression?

No!

$$J = \frac{1}{n}\sum_{i=1}^{n}(pred_i - y_i)^2$$

$$pred_i = \frac{1}{1 + e^{-x_i}}$$

Definition: A function is called convex if the line segment between any two points on the graph lies above the graph between the two points.

Apply a similar format aforementioned into the cost function of logistic regression?

No!

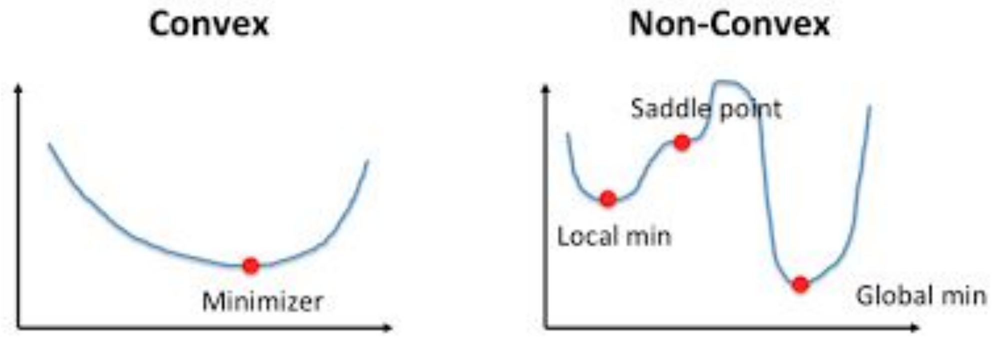
$$J = \frac{1}{n}\sum_{i=1}^{n}(pred_i - y_i)^2$$

$$pred_i = \frac{1}{1+e^{-x_i}}$$

Definition: A function is called convex if the line segment between any two points on the graph lies above the graph between the two points.

Apply a similar format aforementioned into the cost function of logistic regression?

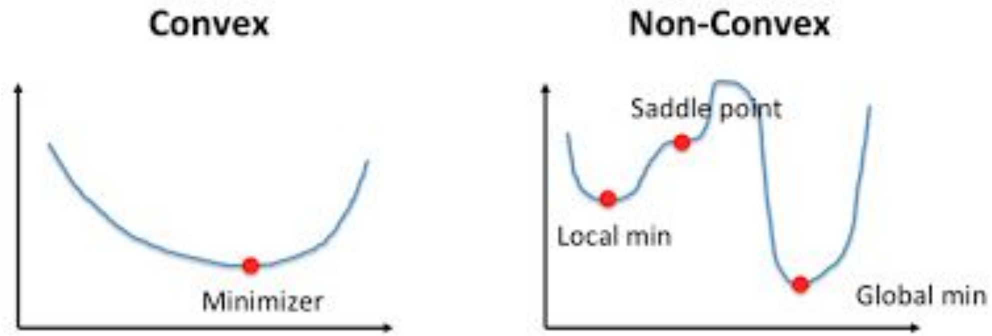
No!



$$J = \frac{1}{n}\sum_{i=1}^{n}(pred_i - y_i)^2$$

$$pred_i = \frac{1}{1 + e^{-x_i}}$$

Definition: A function is called convex if the line segment between any two points on the graph lies above the graph between the two points.



Gradient Descent?

Apply a similar format aforementioned into the cost function of logistic regression?

No!

$$J = \frac{1}{n}\sum_{i=1}^{n}(pred_i - y_i)^2$$

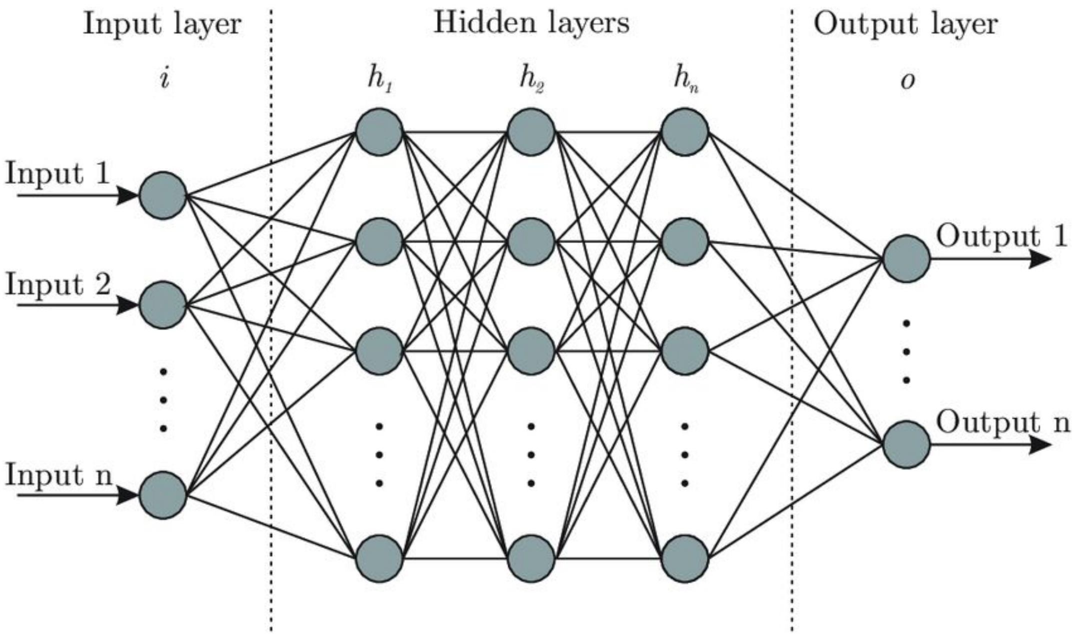
$$pred_i = \frac{1}{1+e^{-x_i}}$$

Cost Function

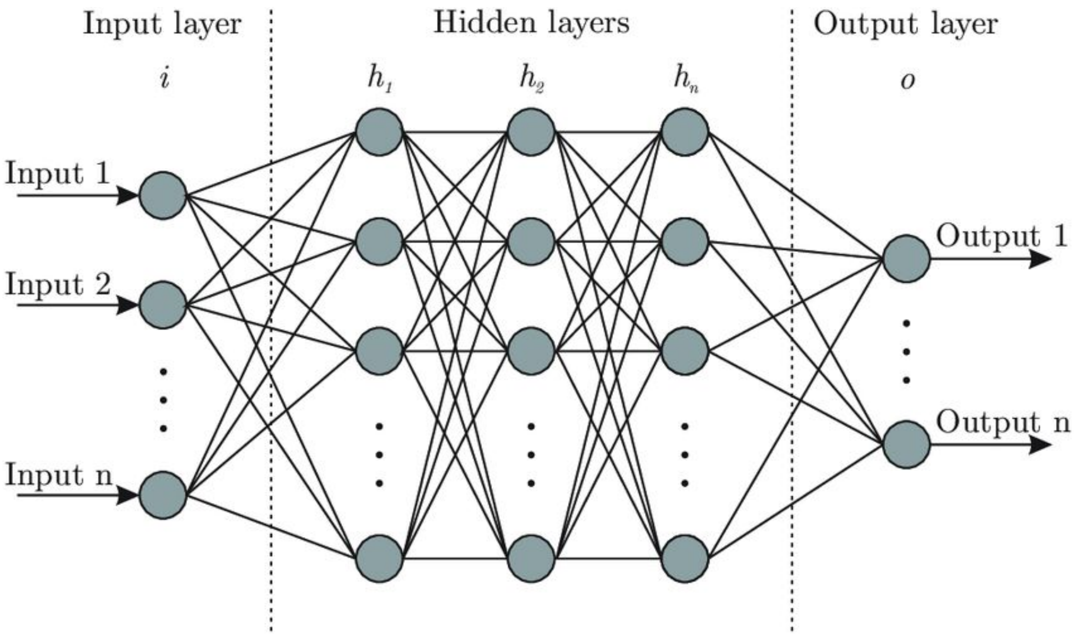
$$g(x) = \frac{1}{1 + e^{-x}}$$

$$Cost(g(x), y) = \begin{cases} -\log(g(x)) & if \;\; y = 1 \\ -\log(1 - g(x)) & if \;\; y = 0 \end{cases}$$

Neural Network Basics



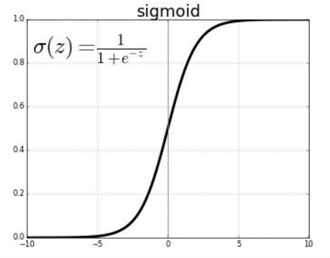
Neural Network Basics

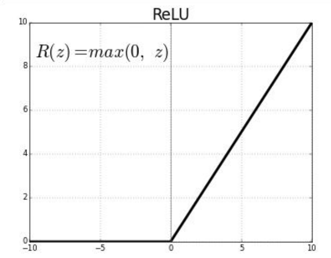


Each Neuron contains a number between 0 and 1, and this number is called **activation**.

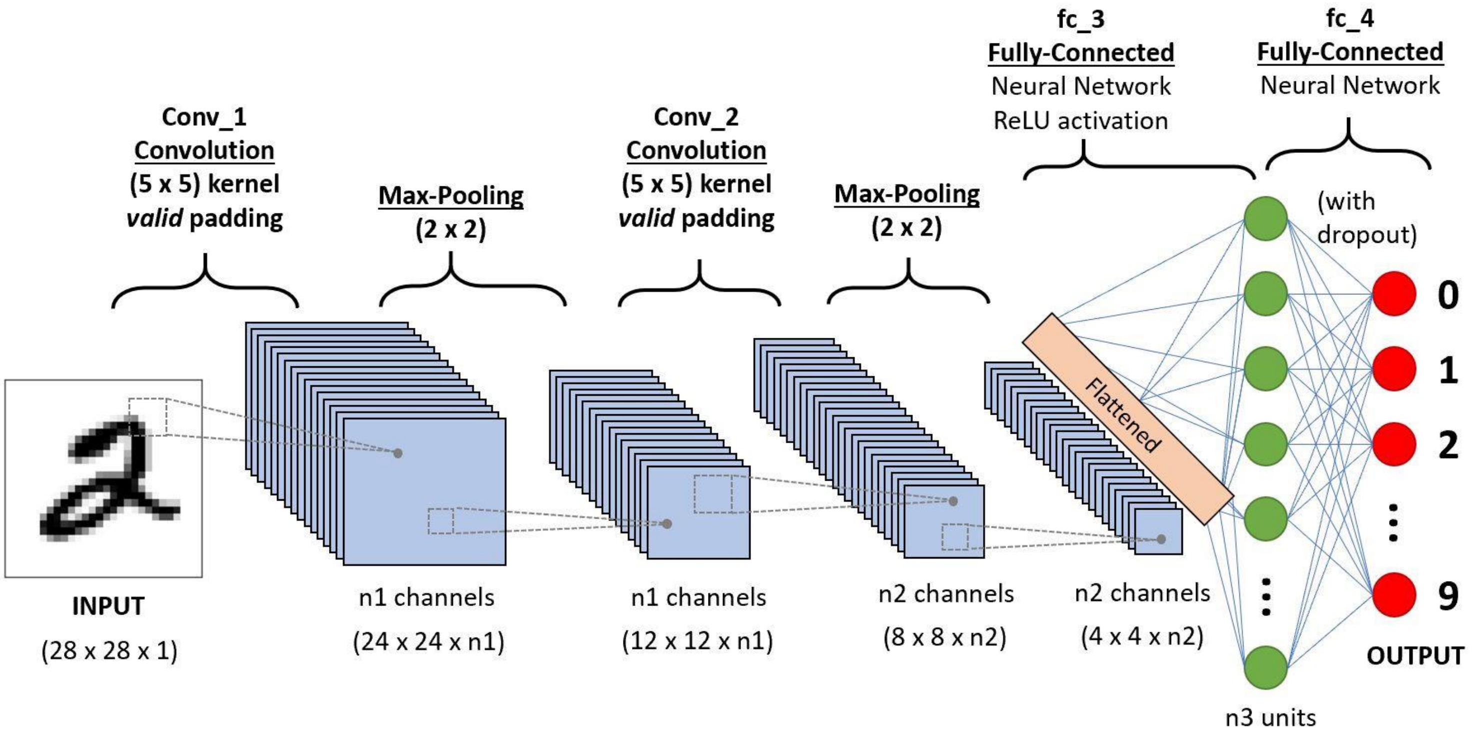
$$\sigma\left(w_1 a_1 + w_2 a_2 + \ldots + w_n a_n - bias \right)$$

Activation function: sigmoid / **ReLU**

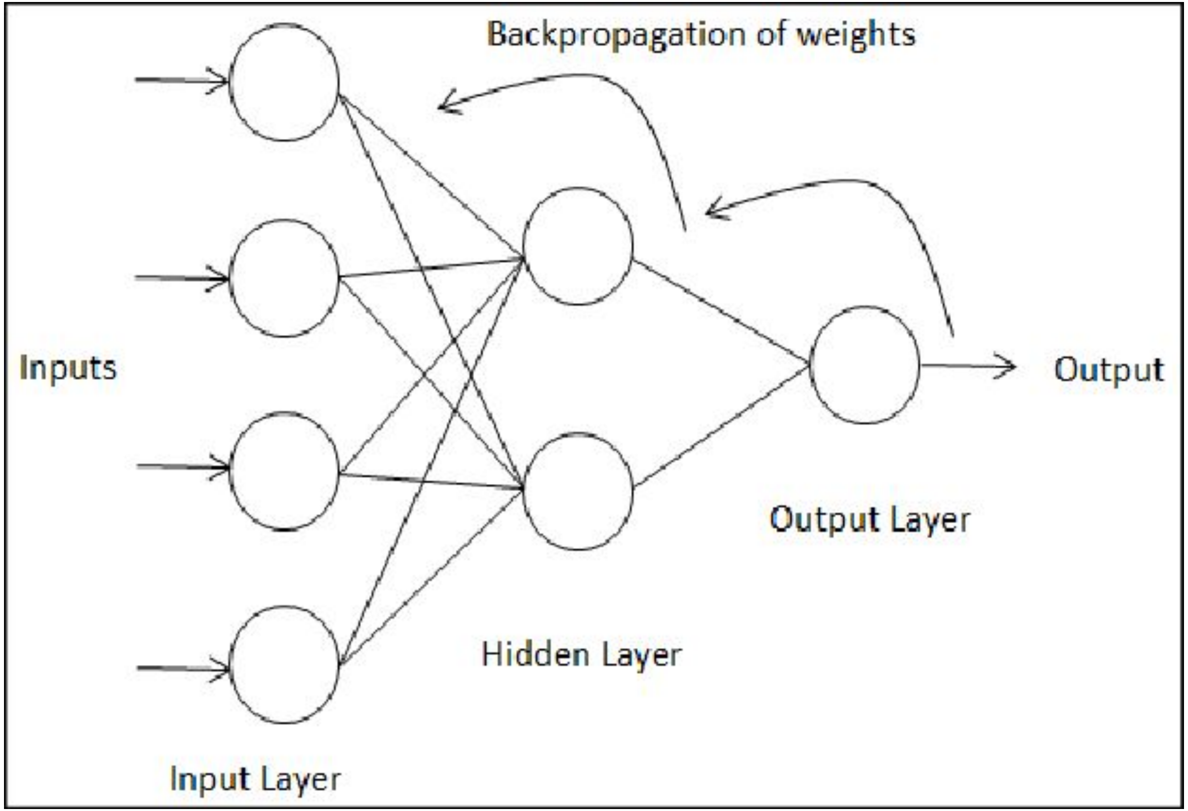




Convolutional Neural Network



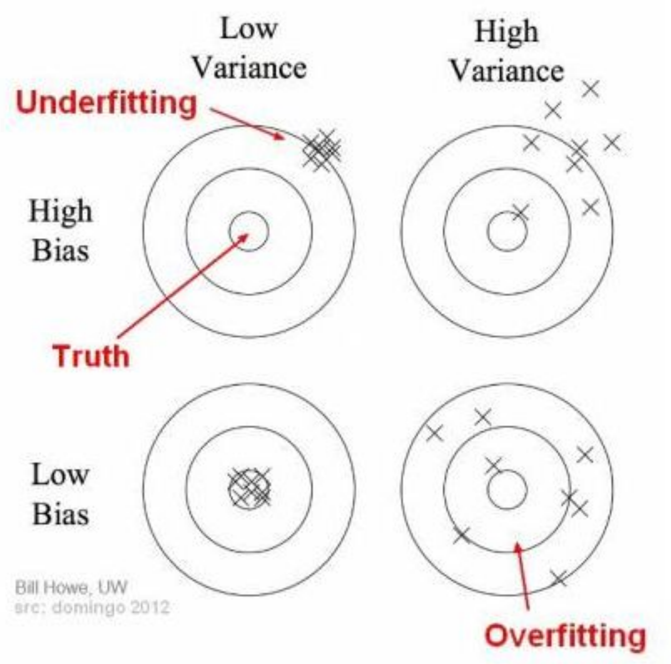
Backpropagation



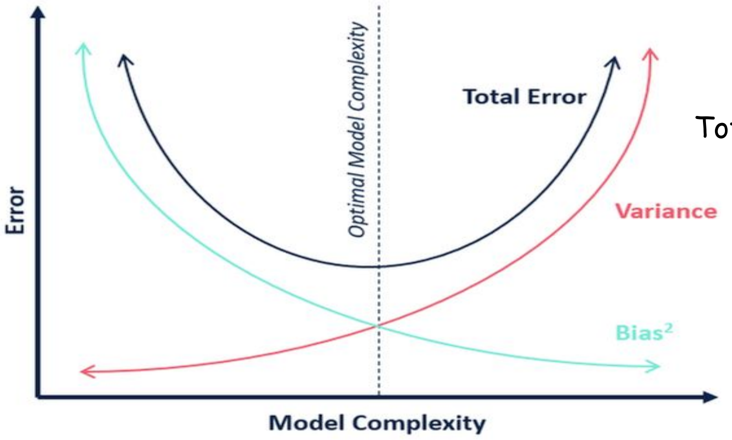
Bias & Variance Tradeoff

Bias: difference between the average model prediction and the correct value

Variance: variability of model prediction for a given data point

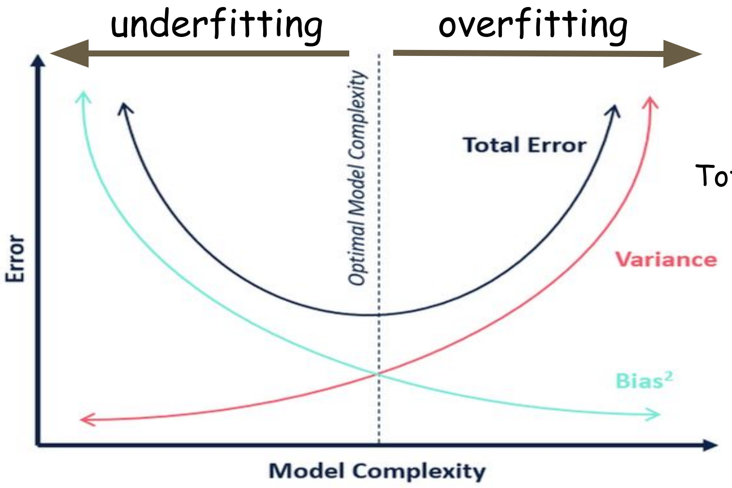


Examples



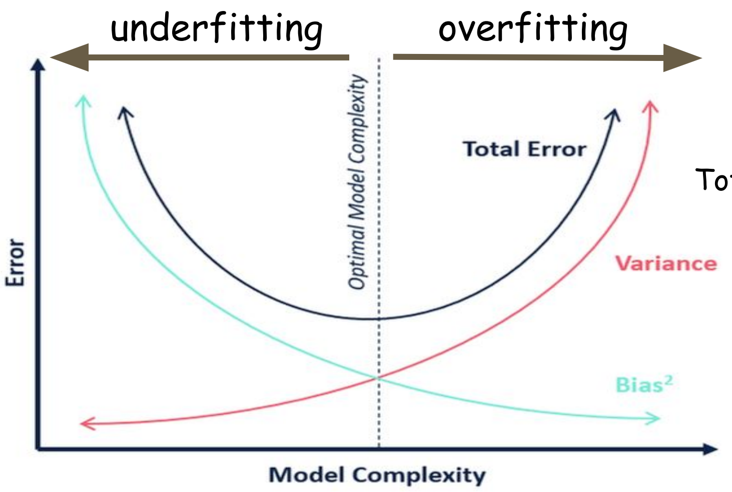
Total Error = Bias^2 + Variance

Examples

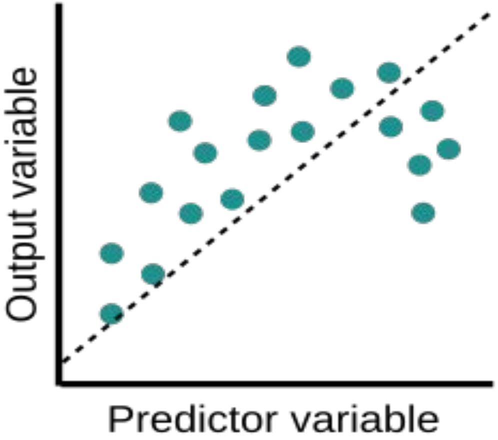


Total Error = Bias^2 + Variance

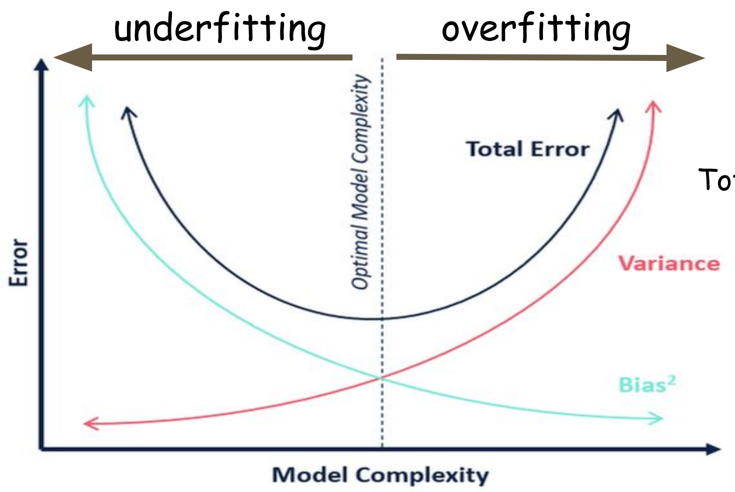
Examples



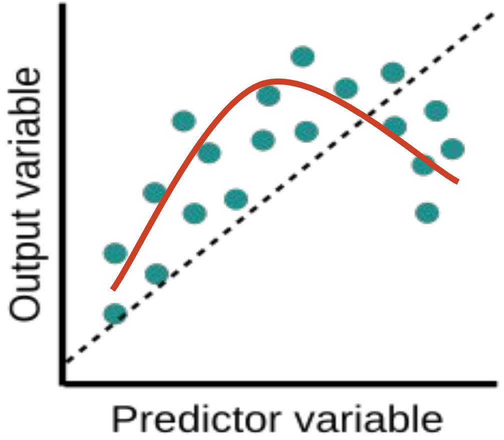
Total Error = Bias^2 + Variance



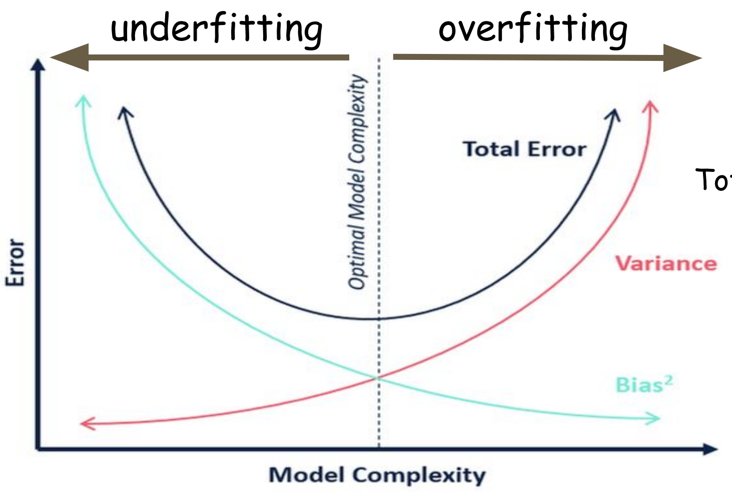
Examples



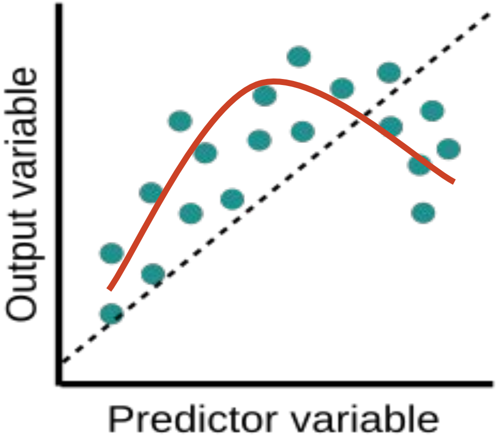
Total Error = Bias^2 + Variance

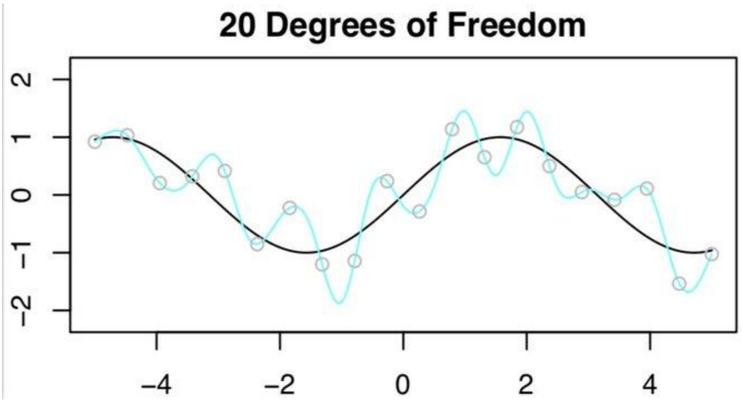


Examples

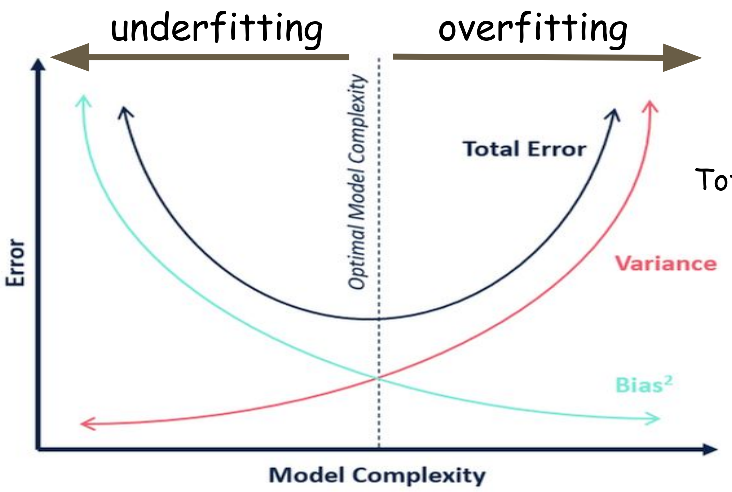


Total Error = Bias^2 + Variance

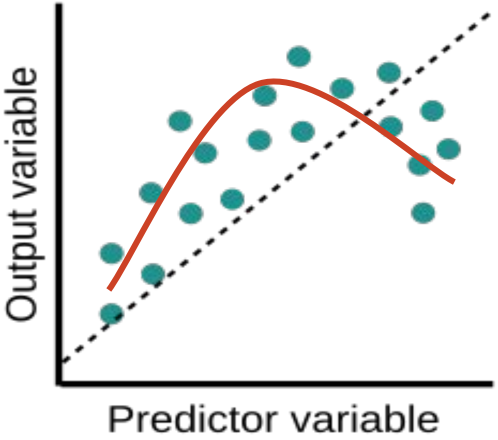


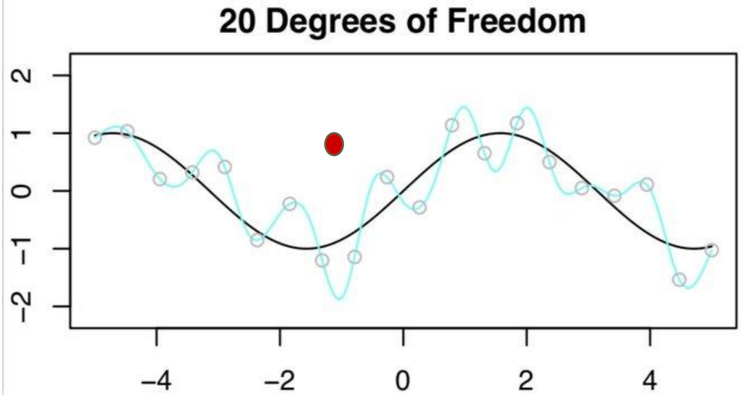


Examples

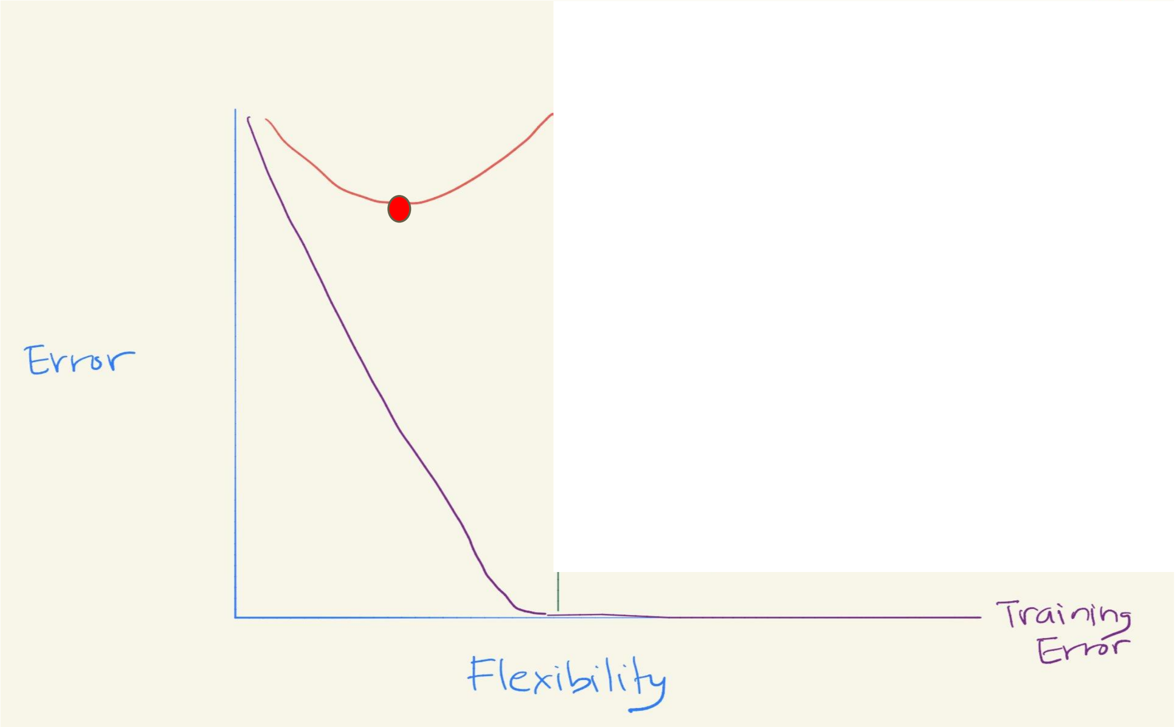


Total Error = Bias^2 + Variance



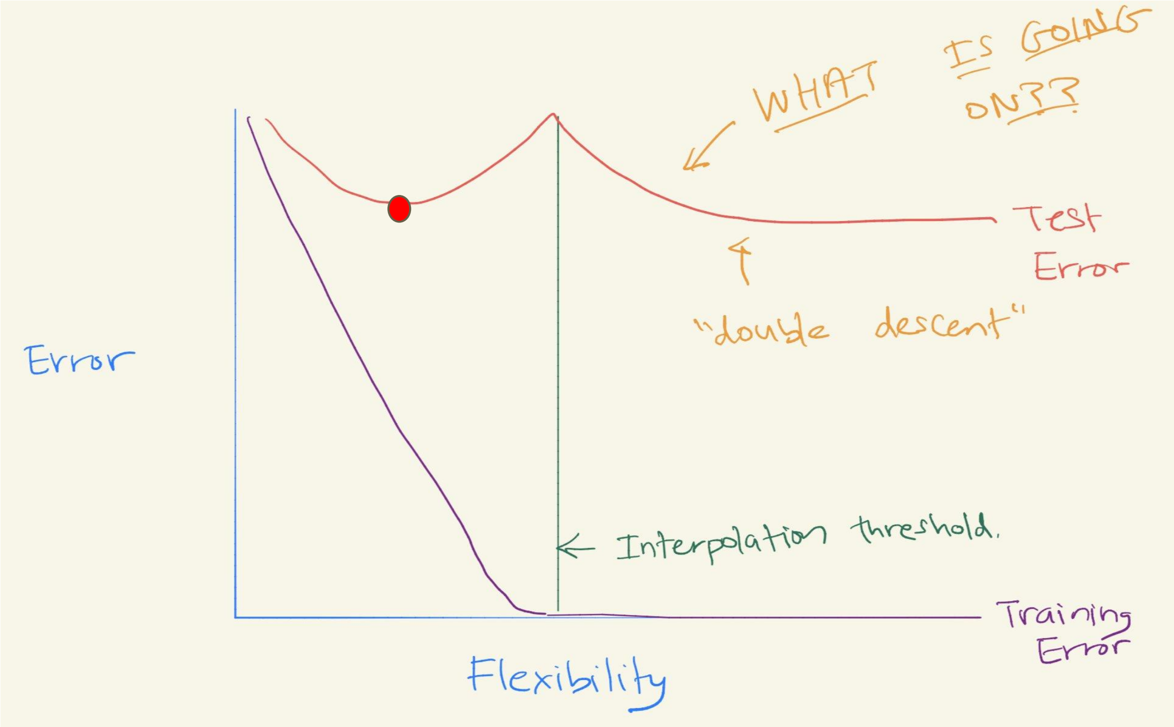


“Magical” Double Descent



From Dr. Daniela Witten, Twitter

“Magical” Double Descent



From Dr. Daniela Witten, Twitter

Python Coding Exercise

```
Epoch 21, Train Accuracy: 0.9193500280380249 , TrainLoss: 0.2311916169751198 , Test Accuracy: 0.7860000133514404
Epoch 22, Train Accuracy: 0.9211500287055969 , TrainLoss: 0.2221628747746403 , Test Accuracy: 0.7680000066757202
Epoch 23, Train Accuracy: 0.93545001745224 , TrainLoss: 0.18967052351204303 , Test Accuracy: 0.777000100135803
Epoch 24, Train Accuracy: 0.9355999827384949 , TrainLoss: 0.1869510811425708 , Test Accuracy: 0.7590000033378601
Epoch 25, Train Accuracy: 0.9390000104904175 , TrainLoss: 0.1745522741011379 , Test Accuracy: 0.765999972820282
Epoch 26, Train Accuracy: 0.9432500004768372 , TrainLoss: 0.16692743151089687 , Test Accuracy: 0.7639999985694885
Epoch 27, Train Accuracy: 0.9429000020027161 , TrainLoss: 0.1596744373470184 , Test Accuracy: 0.746999979019165
Epoch 28, Train Accuracy: 0.9503499865531921 , TrainLoss: 0.14116537495478299 , Test Accuracy: 0.7839999794960022
Epoch 29, Train Accuracy: 0.9498500227928162 , TrainLoss: 0.14538846581803286 , Test Accuracy: 0.7689999938011169
Epoch 30, Train Accuracy: 0.9565500020980835 , TrainLoss: 0.13069106599281055 , Test Accuracy: 0.765999972820282
```

References

- 3Blue1Brown & StatQuest & Andrew Ng YouTube
- Towards Data Science: Logistic Regression Detailed Overview
- Twitter from Dr. Daniela Witten on Bias and Variance Tradeoff
- Heartbeat: Basics of Image Classification

You can click on these 3 articles if interested!

Thanks for Listening!