Module No. 1
Introduction to Machine Learning

Dr. Mrunal Rane

Syllabus :

- **Introduction to Machine Learning:**
  - Types of Machine Learning, Issues in Machine Learning,
  - Application of Machine Learning,
  - Steps involved in developing a Machine Learning Application.
  - **Hypothesis and Inductive Bias**

Supervised Learning

Features (X) | Output (Y)

Instances

User ID	Gender	Age	Salary	Purchased
15624510	Male	19	19000	0
15810944	Male	35	20000	1
15668575	Female	26	43000	0
15603246	Female	27	57000	0
15804002	Male	19	76000	1
15728773	Male	27	58000	1
15598044	Female	27	84000	0
15694829	Female	32	150000	1
15600575	Male	25	33000	1
15727311	Female	35	65000	0
15570769	Female	26	80000	1
15606274	Female	26	52000	0
15746139	Male	20	86000	1
15704987	Male	32	18000	0
15628972	Male	18	82000	0
15697686	Male	29	80000	0
15733883	Male	47	25000	1

Discrete

Features (X) | Output (Y)

Temperature	Pressure	Relative Humidity	Wind Direction	Wind Speed
10.69261758	986.882019	54.19337313	195.7150879	3.278597116
13.59184184	987.8729248	48.0648859	189.2951202	2.909167767
17.70494885	988.1119385	39.11965597	192.9273834	2.973036289
20.95430404	987.8500366	30.66273218	202.0752869	2.965289593
22.9278274	987.2833862	26.06723423	210.6589203	2.798230886
24.04233986	986.2907104	23.46918024	221.1188507	2.627005816
24.41475295	985.2338867	22.25082295	233.7911987	2.448749781
23.93361956	984.8914795	22.35178837	244.3504333	2.454271793
22.68800023	984.8461304	23.7538641	253.0864716	2.418341875
20.56425726	984.8380737	27.07867944	264.5071106	2.318677425
17.76400389	985.4262085	33.54900114	280.7827454	2.343950987
11.25680746	988.9386597	53.74139903	68.15406036	1.650191426
14.37810685	989.6819458	40.70884681	72.62069702	1.553469896
18.45114201	990.2960205	30.85038484	71.70604706	1.005017161
22.54895853	989.9562988	22.81738811	44.66042709	0.264133632
24.23155922	988.796875	19.74790765	318.3214111	0.329656571

Continuous

Inductive

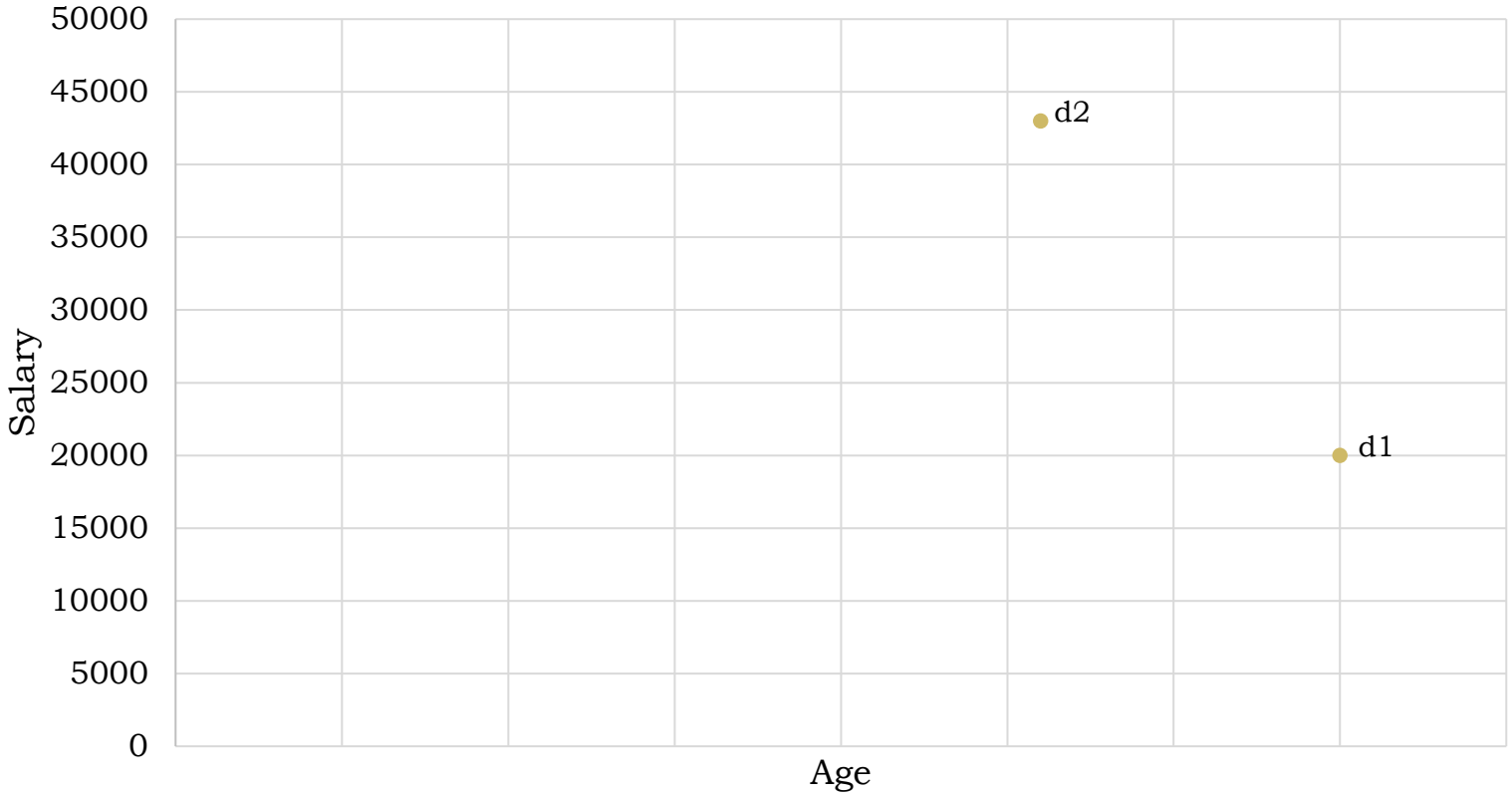
Using specific observations to form a general conclusion

VS

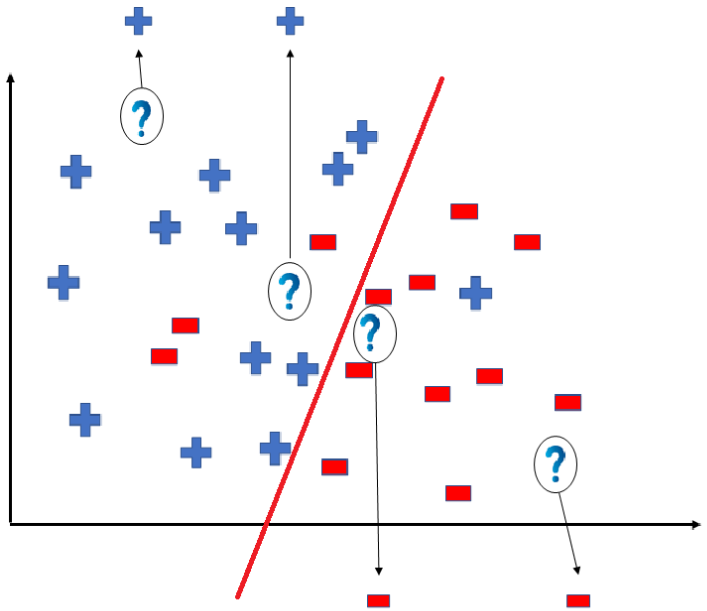
Deductive

Using a general premise to form a specific conclusion

Feature space

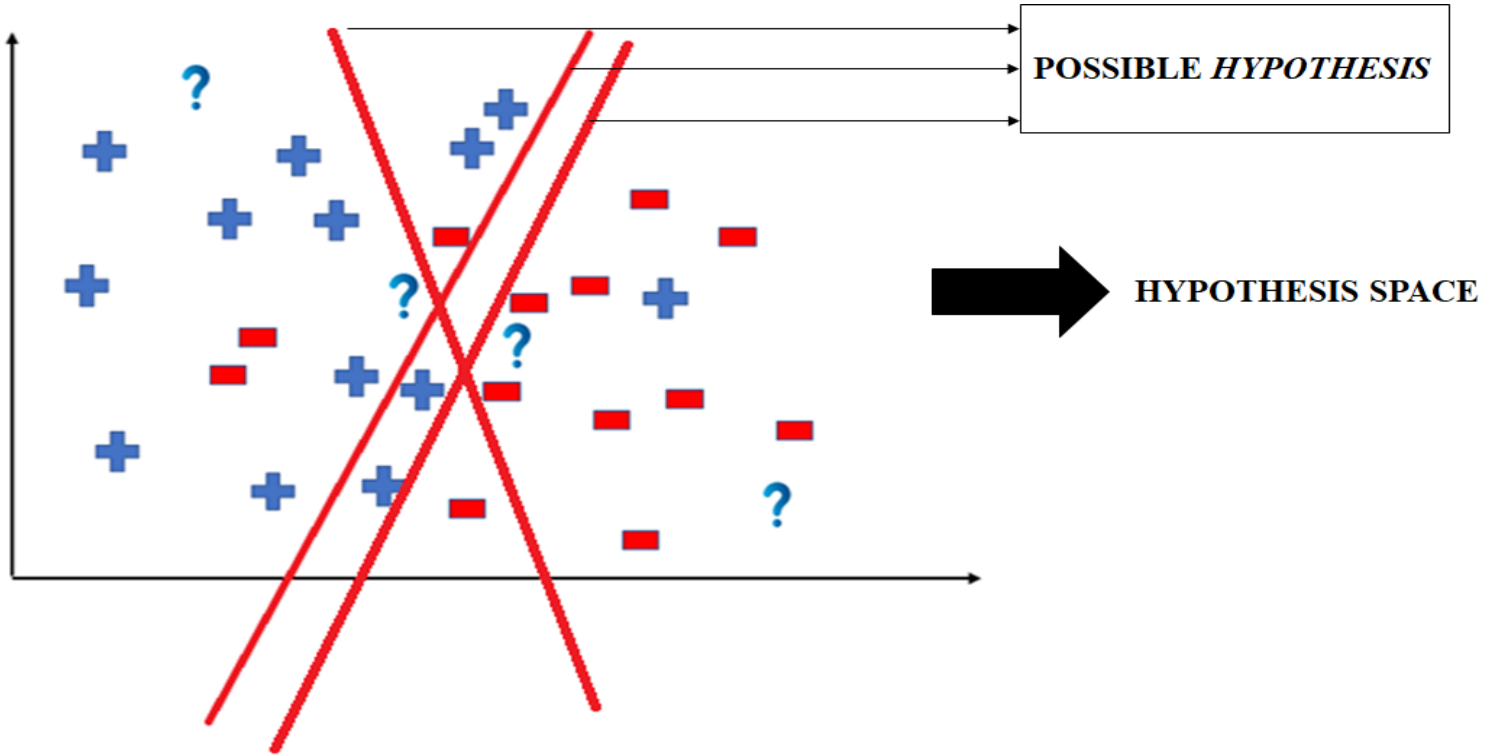


Hypothesis

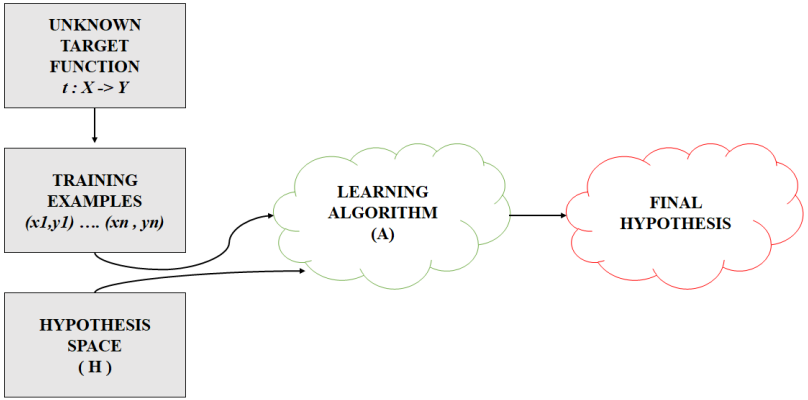


- **Hypothesis (h)**

A hypothesis is a function that best describes the target in supervised machine learning. The hypothesis that an algorithm would come up depends upon the data and also depends upon the restrictions and bias that we have imposed on the data.


POSSIBLE *HYPOTHESIS*
HYPOTHESIS SPACE

Hypothesis



- **Hypothesis Space (H)**

Hypothesis space is the set of all the possible legal hypothesis. This is the set from which the machine learning algorithm would determine the best possible (only one) which would best describe the target function or the outputs.

Data:

X	Y
2	3.4
3	5.9
5	7.8
7.8	6.5
9.2	11.7
10.4	15.3
11.8	17.6

Model? → Hypothesis?

- Data

X	Y
2	3.4
3	5.9
5	7.8
7.8	6.5
9.2	11.7
10.4	15.3
11.8	17.6

- Hypothesis?

1. $Y = bx + a$
2. $Y = e^{-(bx)}$
3. $Y = \text{Sin}(bx)$
4. $Y = bx^2$
5. $Y = \sqrt{a + bx}$

Inductive bias

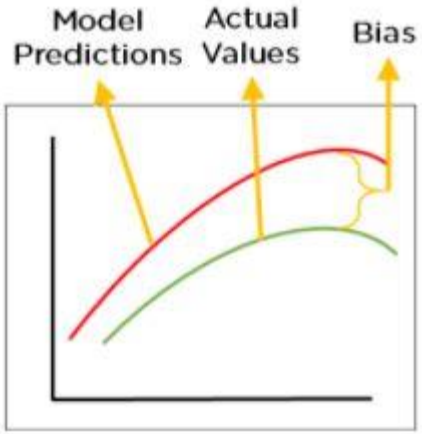
Inductive bias

- The inductive bias (also known as learning bias) of a learning algorithm is the set of assumptions that the learner uses to predict outputs.
- Y = a + b*X

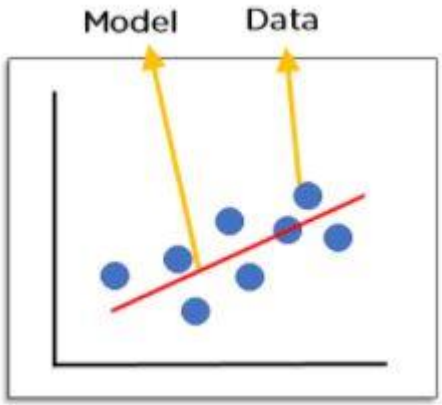
Bias and variance

- used in supervised machine learning
- The correct balance of bias and variance is vital
- bias-variance trade-off

Bias



High bias = underfitting

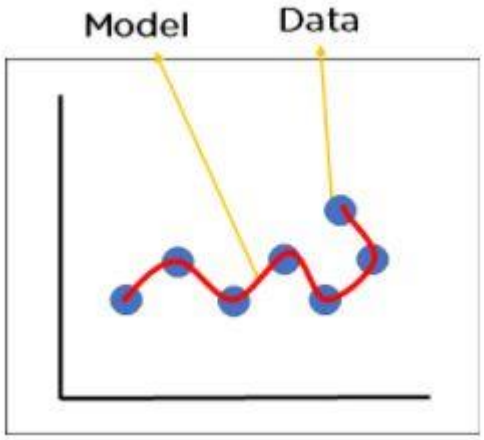


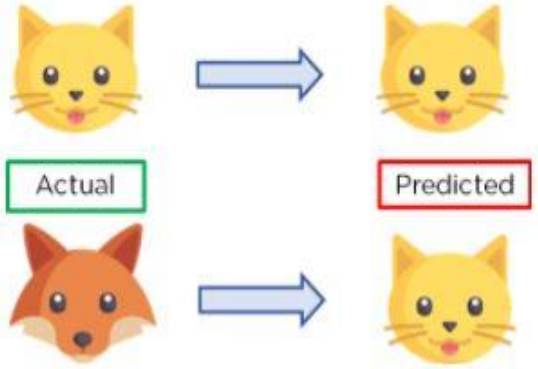
- **Low Bias:**
  - means fewer assumptions are taken to build the target function.
  - the model will closely match the training dataset.
- **High Bias:**
  - means more assumptions are taken to build the target function.
  - the model will not match the training dataset closely.

Variance

High variance = overfitting

- the measure of spread in data from its mean position.
- is the amount by which the performance of a predictive model changes when it is trained on different subsets of the training data.
- is the variability of the model that how much it is sensitive to another subset of the training dataset. i.e. how much it can adjust on the new subset of the training dataset.



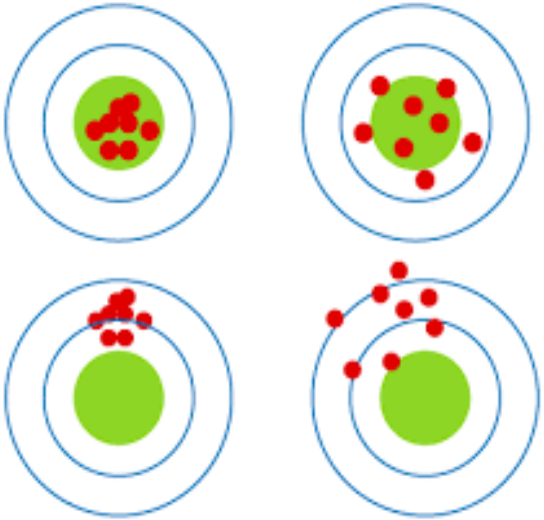


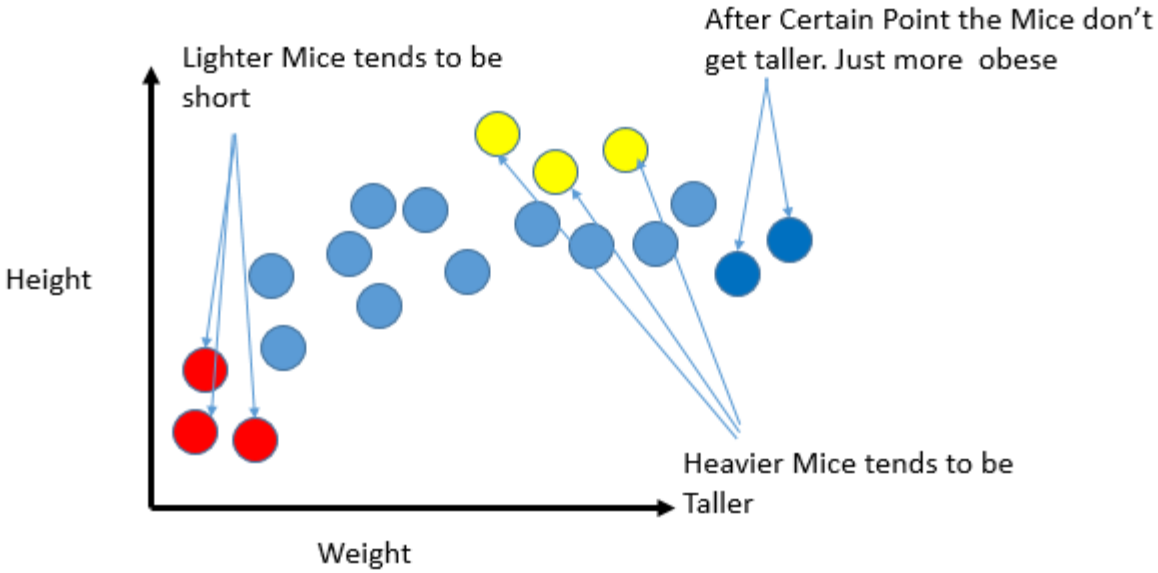
- **Low variance:**
  - means that the model is less sensitive to changes in the training data and can produce consistent estimates of the target function with different subsets of data from the same distribution.
  - This is the case of underfitting when the model fails to generalize on both training and test data.

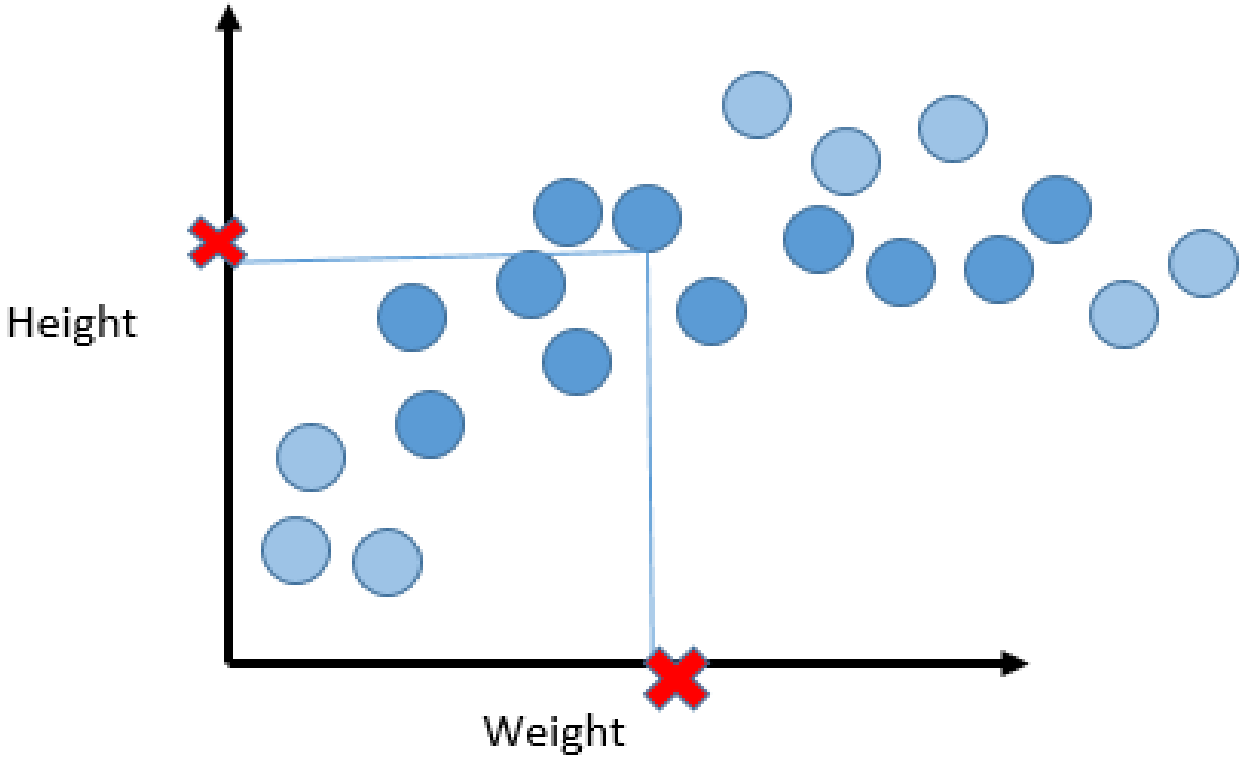
- **High variance:**
  - means that the model is very sensitive to changes in the training data and can result in significant changes in the estimate of the target function when trained on different subsets of data from the same distribution.
  - This is the case of overfitting when the model performs well on the training data but poorly on new, unseen test data. It fits the training data too closely that it fails on the new training dataset.

Combinations of Bias-Variance :

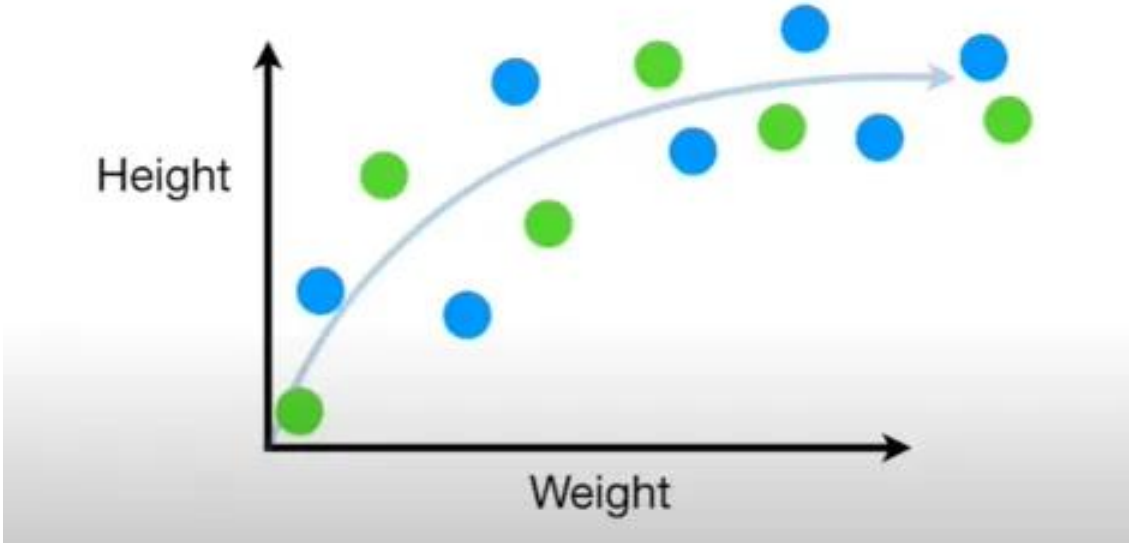
- ***Low-Bias, Low-Variance:*** The combination of low bias and low variance shows an ideal machine learning model. However, it is not possible practically.
- ***Low-Bias, High-Variance:*** With low bias and high variance, model predictions are inconsistent and accurate on average. This case occurs when the model learns with a large number of parameters and hence leads to an **overfitting**
- ***High-Bias, Low-Variance:*** With High bias and low variance, predictions are consistent but inaccurate on average. This case occurs when a model does not learn well with the training dataset or uses few numbers of the parameter. It leads to **underfitting** problems in the model.
- ***High-Bias, High-Variance:*** With high bias and high variance, predictions are inconsistent and also inaccurate on average.

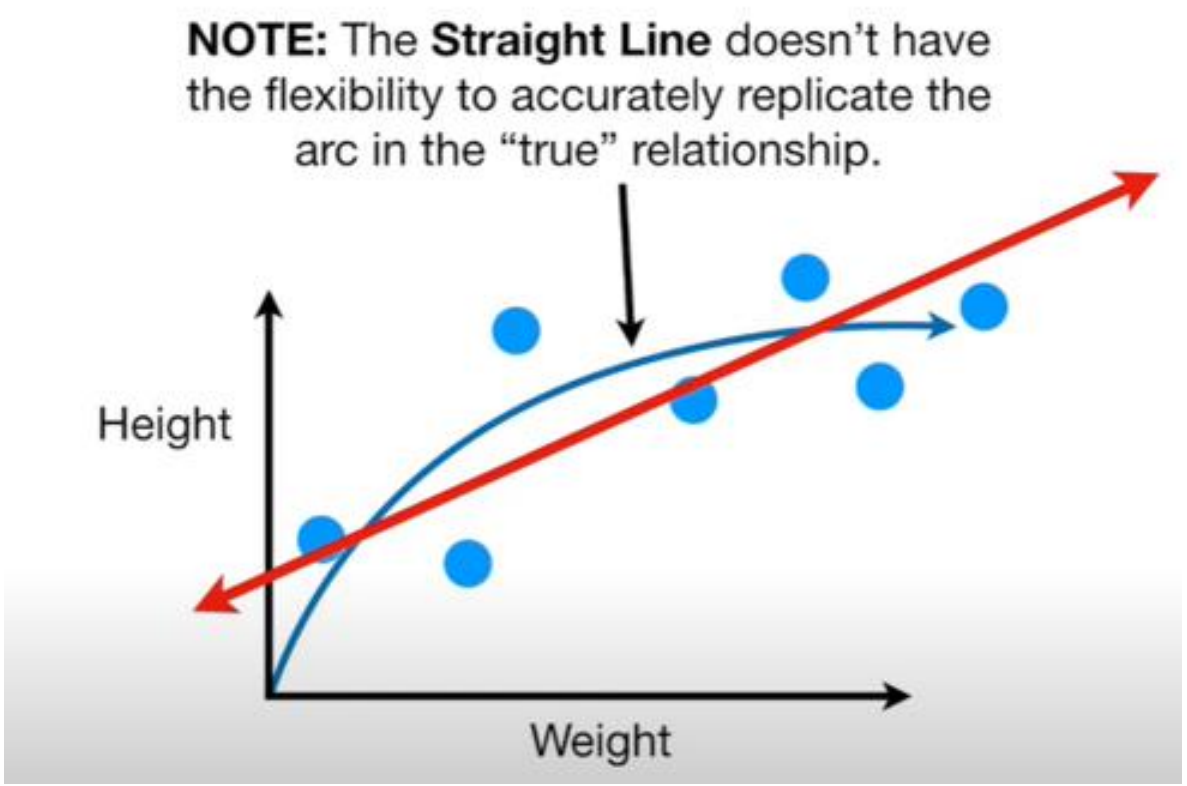



After Certain Point the Mice don't get taller. Just more obese
Lighter Mice tends to be short
Height
Heavier Mice tends to be Taller
Weight

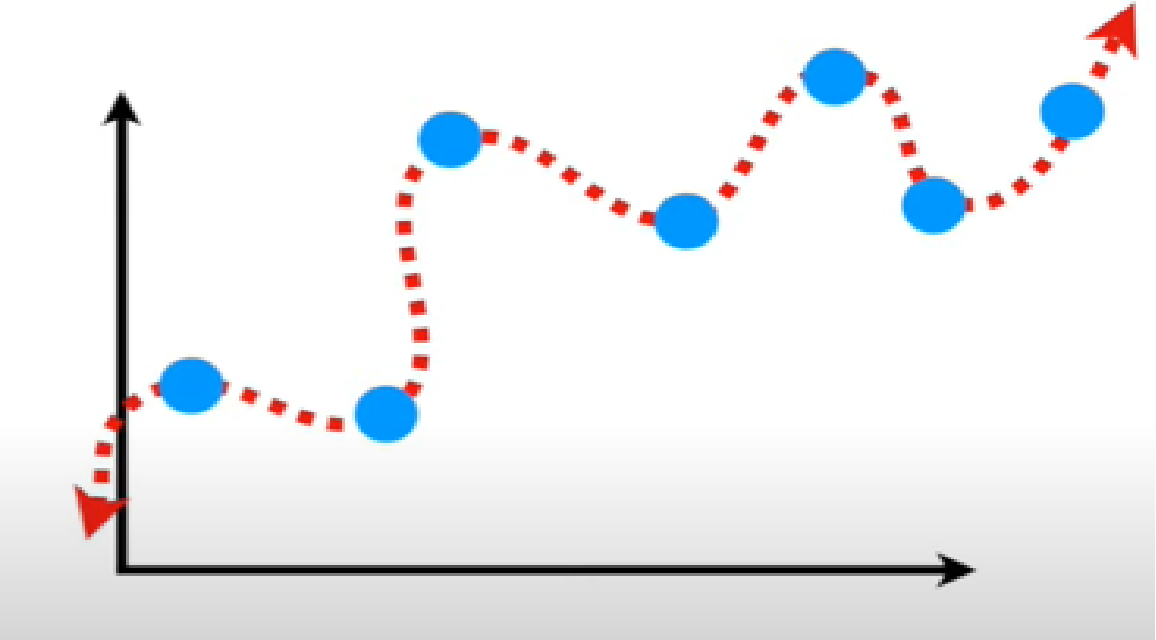

Height
Weight

The first thing we do is split the data into two sets, one for training the machine learning algorithms and one for testing them.

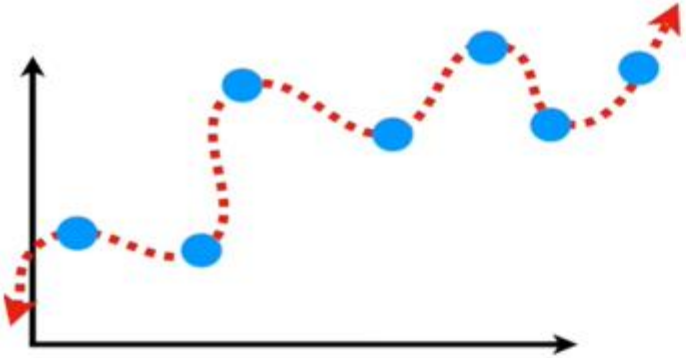



NOTE: The Straight Line doesn't have the flexibility to accurately replicate the arc in the "true" relationship.
Height
Weight

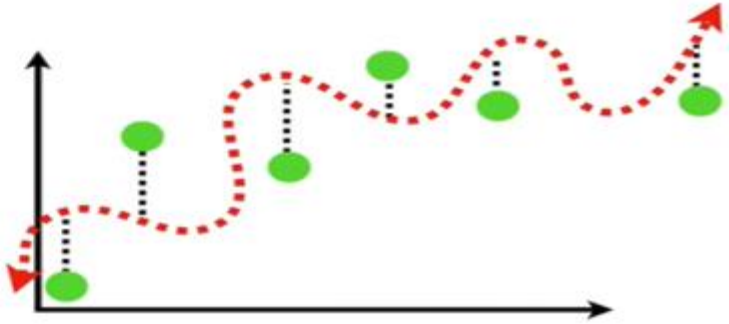
Notice how the **Squiggly Line** fits the data so well that the distances between the line and the data are all 0.



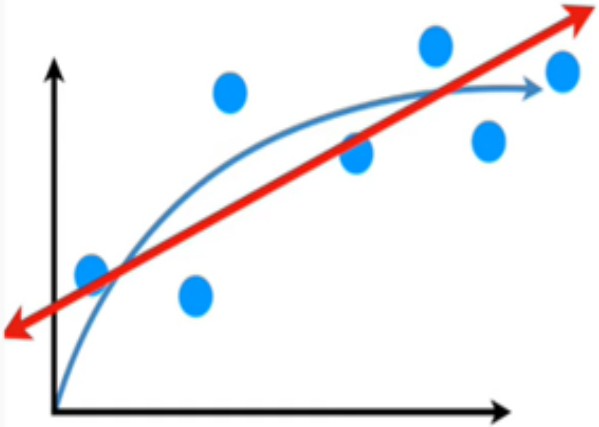
Even though **Squiggly Line** did a great job fitting the **training set**...



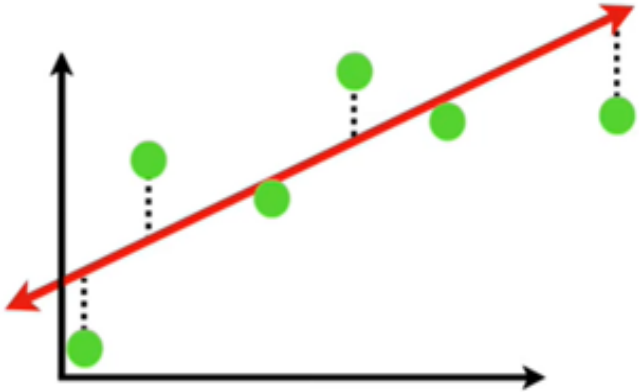
...it did a terrible job fitting the **testing set**...



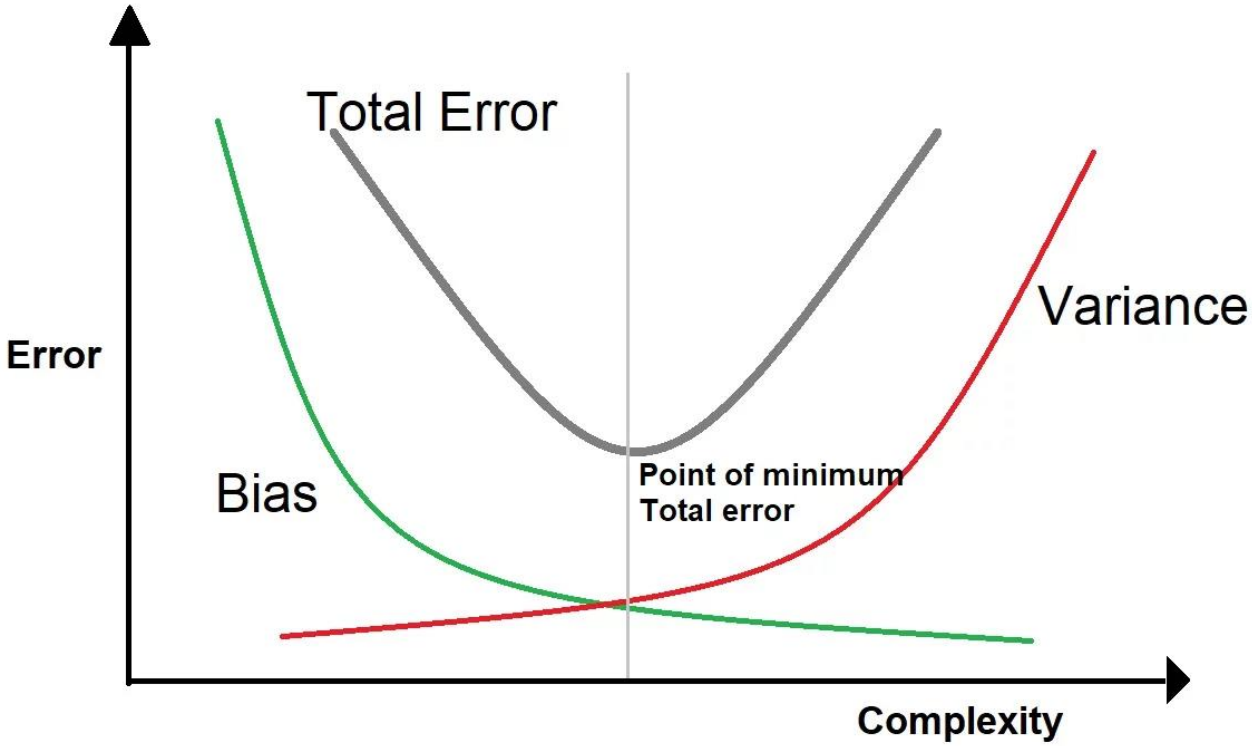
In contrast, the **Straight Line** has relatively **high bias**, since it can not capture the curve in the relationship between weight and height...



...but the **Straight Line** has relatively **low variance**, because the Sums of Squares are very similar for different datasets.



Trade-off


Total Error
Variance
Error
Bias
Point of minimum
Total error
Complexity