Naive Bayes

→ Naive Bayes usually deals with independent features of a dataset.

→ Suppose in previous example we are predicting whether a car is cheap & expensive.

→ Let's expand our features using Rules of Probability along with Bayes's rule

$$P(X_1 X_2 X_3 | Y) = P(X_1) \cdot P(X_1|Y) \cdot P(X_2|X_1, Y) \cdot P(X_3|X_2 X_1, Y)$$

This is derived from chain rule of probability.

→ Now Let's assume the our features are Conditionally independent. So for them taking the additional feature of X_1 or $X_2 X_1$ doesn't really matter. So expression would be

$$P(X_1 X_2 X_3 | Y) = P(Y) \cdot P(X_1|Y) \cdot P(X_2|Y) \cdot P(X_3|Y)$$

→ Now for example if we want to find if Honda Car's are Costly or cheap. Here the feature is "Manufacturer"

→ Go through the dataset and look how many Honda cars are cheap.
Suppose let it be 10.
So total no of cheap cars are 25
then $P(X_1 = \text{Honda} | Y = \text{cheap}) = \frac{10}{25} = \frac{2}{5}$

-> Like wise we can compute for other features also. Then using all the features we can multiply their probabilities and find out whether it is cheap or not. This can be done for expensive case also.

-> Naive bayes can be used in spam filtering, determing a specific author's book.

-> The issue comes when the data (features) are more correleted between each other.
For example if we are classifying a email message as positive & negative sentiment and we use features for whether or not a message contains certain words.

-> When the features are correlated then there would be a problem and our model would underperform