`START with this (for concept): https://www.youtube.com/watch?v=GBMMtXRiQX0`

Step 1: Define the Features and Classes

The **Iris dataset** has:

- Features: x_1 = sepal length, x_2 = sepal width, x_3 = petal length, x_4 = petal width
- Classes (species): C_1 = Setosa, C_2 = Versicolor, C_3 = Virginica

Given a flower with feature vector (x_1, x_2, x_3, x_4), we aim to classify it into one of the three species.

Step 2: Bayes Theorem

For Naive Bayes, we calculate the posterior probability of each class C_i given the features (x_1, x_2, x_3, x_4):

$$P(C_i|x_1, x_2, x_3, x_4) = \frac{P(C_i) \cdot P(x_1|C_i) \cdot P(x_2|C_i) \cdot P(x_3|C_i) \cdot P(x_4|C_i)}{P(x_1, x_2, x_3, x_4)}$$

We ignore the denominator $P(x_1, x_2, x_3, x_4)$ since it is the same for all classes. Thus, we only need to compute the numerator for each class:

$$P(C_i|x_1, x_2, x_3, x_4) \propto P(C_i) \cdot P(x_1|C_i) \cdot P(x_2|C_i) \cdot P(x_3|C_i) \cdot P(x_4|C_i)$$

Step 3: Prior Probabilities $P(C_i)$

The **prior probability** $P(C_i)$ is the proportion of each class in the dataset. Assuming the Iris dataset has an equal number of samples for each class (50 samples per class out of 150):

$$P(C_1) = P(\text{Setosa}) = \frac{50}{150} = 0.33$$

$$P(C_2) = P(\text{Versicolor}) = \frac{50}{150} = 0.33$$

$$P(C_3) = P(\text{Virginica}) = \frac{50}{150} = 0.33$$

Step 4: Likelihood $P(x_j|C_i)$

The likelihood $P(x_j|C_i)$ is calculated based on the feature distributions (mean and variance) for each class. Typically, Naive Bayes assumes a **Gaussian distribution** for continuous features:

$$P(x_j|C_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x_j - \mu_i)^2}{2\sigma_i^2}\right)$$

Where:

- μ_i is the mean of the feature for class C_i,
- σ_i^2 is the variance of the feature for class C_i,
- x_j is the feature value (e.g., the sepal length of the test flower).

You would compute this for each feature (sepal length, sepal width, petal length, petal width) and for each class (Setosa, Versicolor, Virginica).

Step 5: Example

Assume you have a flower with these feature values:

- Sepal length = 5.1 cm
- Sepal width = 3.5 cm
- Petal length = 1.4 cm
- Petal width = 0.2 cm

And let's say we have the following Gaussian parameters for each class:

Class	μ_1 **(sepal length)**	σ_1^2	μ_2 **(sepal width)**	σ_2^2	μ_3 **(petal length)**	σ_3^2	μ_4 **(petal width)**	σ_4^2
Setosa	5.01	0.12	3.42	0.14	1.46	0.03	0.24	0.01
Versicolor	5.94	0.26	2.77	0.09	4.26	0.22	1.33	0.04
Virginica	6.59	0.30	2.97	0.12	5.55	0.27	2.03	0.05

For each feature, calculate the likelihood $P(x_j|C_i)$ using the Gaussian distribution formula for each class.

For Setosa:

$$P(x_1 = 5.1|\text{Setosa}) = \frac{1}{\sqrt{2\pi \cdot 0.12}} \exp\left(-\frac{(5.1 - 5.01)^2}{2 \cdot 0.12}\right)$$

$$P(x_2 = 3.5|\text{Setosa}) = \frac{1}{\sqrt{2\pi \cdot 0.14}} \exp\left(-\frac{(3.5 - 3.42)^2}{2 \cdot 0.14}\right)$$

And so on for petal length and width.

For Versicolor:

$$P(x_1 = 5.1|\text{Versicolor}) = \frac{1}{\sqrt{2\pi \cdot 0.26}} \exp\left(-\frac{(5.1 - 5.94)^2}{2 \cdot 0.26}\right)$$

$$P(x_2 = 3.5|\text{Versicolor}) = \frac{1}{\sqrt{2\pi \cdot 0.09}} \exp\left(-\frac{(3.5 - 2.77)^2}{2 \cdot 0.09}\right)$$

And similarly for the remaining features and classes.

Step 6: Posterior Calculation

After computing all the likelihoods and multiplying by the priors, you will have three posterior probabilities, one for each class.

Step 7: Prediction

Choose the class with the highest posterior probability. If the posterior for Setosa is highest, then the flower is classified as **Setosa**.

If we avoid using the **Gaussian distribution** (which is typically applied for continuous features like in the Iris dataset), we can still apply Naive Bayes in a **non-parametric** manner.

Problem Setup:

We have 4 features in the Iris dataset:

- x_1: sepal length
- x_2: sepal width
- x_3: petal length
- x_4: petal width

The goal is to classify a flower into one of the three classes: **Setosa**, **Versicolor**, or **Virginica**, based on its features.

Step 1: Bayes Theorem without Gaussian Distribution

For each class C_i (Setosa, Versicolor, Virginica), we want to calculate the posterior probability:

$$P(C_i|x_1, x_2, x_3, x_4) \propto P(C_i) \cdot P(x_1|C_i) \cdot P(x_2|C_i) \cdot P(x_3|C_i) \cdot P(x_4|C_i)$$

Step 2: Prior Probability $P(C_i)$

The prior probabilities are the proportions of each class in the dataset. For the Iris dataset, with 50 samples per class out of 150 total samples, we have:

$$P(\text{Setosa}) = P(C_1) = \frac{50}{150} = 0.33$$

$$P(\text{Versicolor}) = P(C_2) = \frac{50}{150} = 0.33$$

$$P(\text{Virginica}) = P(C_3) = \frac{50}{150} = 0.33$$

Step 3: Likelihood $P(x_j|C_i)$

Since we are not using Gaussian distributions for the features, we will calculate the likelihood $P(x_j|C_i)$ directly from the dataset based on **frequency counts**. In this case, we discretize the feature values into **bins** or **ranges** and calculate the likelihoods based on how often each value or range appears for a given class.

Example of Discretizing Sepal Length:

For simplicity, let's assume we create **bins** for sepal length, such as:

- Bin 1: $[4.0, 5.0]$
- Bin 2: $[5.0, 6.0]$
- Bin 3: $[6.0, 7.0]$
- Bin 4: $[7.0, 8.0]$

We count how many times a sepal length value falls into each bin for each class (Setosa, Versicolor, Virginica).

Likelihood Calculation for Sepal Length:

- For $x_1 =$ sepal length, assume the test flower has a sepal length of **5.1**. This would fall into **Bin 2** ($[5.0, 6.0]$).

Suppose in the dataset we observe:

- Setosa: 30 flowers fall in Bin 2.
- Versicolor: 15 flowers fall in Bin 2.
- Virginica: 5 flowers fall in Bin 2.

Thus, the likelihoods for the sepal length $x_1 = 5.1$ are:

$$P(x_1 = 5.1|\text{Setosa}) = \frac{30}{50} = 0.6$$

$$P(x_1 = 5.1|\text{Versicolor}) = \frac{15}{50} = 0.3$$

$$P(x_1 = 5.1|\text{Virginica}) = \frac{5}{50} = 0.1$$

Repeat for Other Features:

Similarly, you discretize the other features (sepal width, petal length, petal width) and calculate their likelihoods by counting occurrences in the bins for each class.

Step 4: Posterior Probability

Now, for each class, multiply the prior probability by the likelihoods of all the features.

For Setosa:

$$P(\text{Setosa}|x_1 = 5.1, x_2, x_3, x_4) \propto P(\text{Setosa}) \cdot P(x_1|\text{Setosa}) \cdot P(x_2|\text{Setosa}) \cdot P(x_3|\text{Setosa}) \cdot P(x_4|\text{Setosa})$$

Substituting the values:

$$P(\text{Setosa}|x_1 = 5.1, ...) \propto 0.33 \cdot 0.6 \cdot P(x_2|\text{Setosa}) \cdot P(x_3|\text{Setosa}) \cdot P(x_4|\text{Setosa})$$

For Versicolor:

$$P(\text{Versicolor}|x_1 = 5.1, x_2, x_3, x_4) \propto 0.33 \cdot 0.3 \cdot P(x_2|\text{Versicolor}) \cdot P(x_3|\text{Versicolor}) \cdot P(x_4|\text{Versicolor})$$

For Virginica:

$$P(\text{Virginica}|x_1 = 5.1, x_2, x_3, x_4) \propto 0.33 \cdot 0.1 \cdot P(x_2|\text{Virginica}) \cdot P(x_3|\text{Virginica}) \cdot P(x_4|\text{Virginica})$$

Step 5: Prediction

Finally, compare the posterior probabilities for each class. The class with the highest posterior probability is the predicted class for the flower.

For example, if the posterior for Setosa is highest, the flower would be classified as **Setosa**.