Naive-Bayes Classifier.

→ Bayes theorem ⇒ $\boxed{P(Y|X) = \frac{P(X|Y) \cdot P(Y)}{P(X)}}$

→ We have to find, $P(Y=y \mid X=(x_1, x_2 .. x_n))$ is maximum for which value of y.

→ $$\underbrace{P(Y|X)}_{\text{Posterior}} = \frac{\overbrace{P(X|Y)}^{\text{Likelihood}} \times \overbrace{P(Y)}^{\text{Prior}}}{\underbrace{P(X)}_{\text{Evidence}}}$$

→ Let's consider X to be categorical variables.

Steps:

(1). Calculate the priors

(2) Calculate likelihoods.

(3) Compute numerator for each y classes (Ignore denominator as it is constant)

(4) Assume all features are independent of each other.

$\therefore$ $P(X=(0,2) \mid Y=1) = P(X_1=0 \mid Y=1) * P(X_2=2 \mid Y=1)$

NOTE { This assumption is not always true. This is why this classifier is called "Naive-Bayes" classifier. Else everything will be 0 for combinations that donot exist.

(5) Muliply likelihood & priors

(6) Normalize to 1

(7) One with highest value is the prediction.

Dealing with continuous features

1. Discretization - Break into intervals.
2. Fit a known Distribution

$$P(X = (x_1, x_2, \dots x_n) \mid Y = y) = \prod f(X_i = x_i \mid Y = y)$$

$P(X = (x_1, x_2, \dots x_n) \mid Y = y)$ → p.d.f. of known distribution

$\prod f(X_i = x_i \mid Y = y)$: features are still independent