

1 NAIVE BAYES CLASSIFICATION DESIGN

$$1a) P(X|Y) = \frac{P(X \cup Y)}{P(Y)} = \frac{P(X \cup Y)}{P(X)} = P(Y|X)$$

$$P(X|Y)P(Y) = P(Y|X)P(X)$$

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

1b) This equation is always true and makes no assumptions on Y , X only that $P(X)$ and $P(Y) > 0$.

$$1c) P(Y | X_1, X_2, \dots, X_{|W|}) = \frac{P(X_1, X_2, \dots, X_{|W|} | Y)P(Y)}{P(X_1, X_2, \dots, X_{|W|})}$$

1d) $P(Y)$ is the probability of decade C_i before X is observed.

$$1e) P(X_1, X_2, \dots, X_{|W|} | Y) = P(X_1 | Y)P(X_2, \dots, X_n | Y, X_1)$$

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

$$= P(X_1 | Y)P(X_2 | Y, X_1) \dots P(X_{|W|} | Y, X_1, X_2, X_3, \dots, X_{n-1})$$

1f)

1g)

1h) The Naive Bayes assumptions means that the features are independent given the class, i.e. assumes conditional independence between X_i and X_j for $i \neq j$ given Y . It is violated when

$$1i) P(Y | X_1, X_2, \dots, X_{|W|}) = \frac{P(Y)P(X_1|Y)P(X_2|Y)P(X_3|Y) \dots P(X_n|Y)}{P(X_1, X_2, \dots, X_{|W|})}$$

$$P(Y | X_1, X_2, \dots, X_{|W|}) = \frac{P(Y) \prod_{i=1}^n P(X_i|Y)}{P(X_1, X_2, \dots, X_{|W|})}$$

2 THE MOMENT OF TRUTH

1a)

1b) The underlying difference between them is in the assumptions they make is in regards to the distribution of the likelihood probabilities $P(X_i | Y)$. Under Gaussian NB, the likelihoods are assumed to be a Gaussian distribution. Multinomial NB assumes a multinomial distribution i.e. that X . And finally Bernoulli NB assumes that each feature has a Bernoulli distribution, or that multiple features only take values 0 or 1. The main difference between Bernoulli and Multinomial is that when calculating the probabilities Multinomial distributions ignore that a particular feature i had no count while Bernoulli would associate a penalty with that.

The closest one for us would be Multinomial,

Our classifier differs from all of these in that it makes no assumptions at all in the distribution of the data, or rather, it gives equal weight to each likelihood.