

with Dr. Mahdi Roozbahani & Wafa Louhichi



G

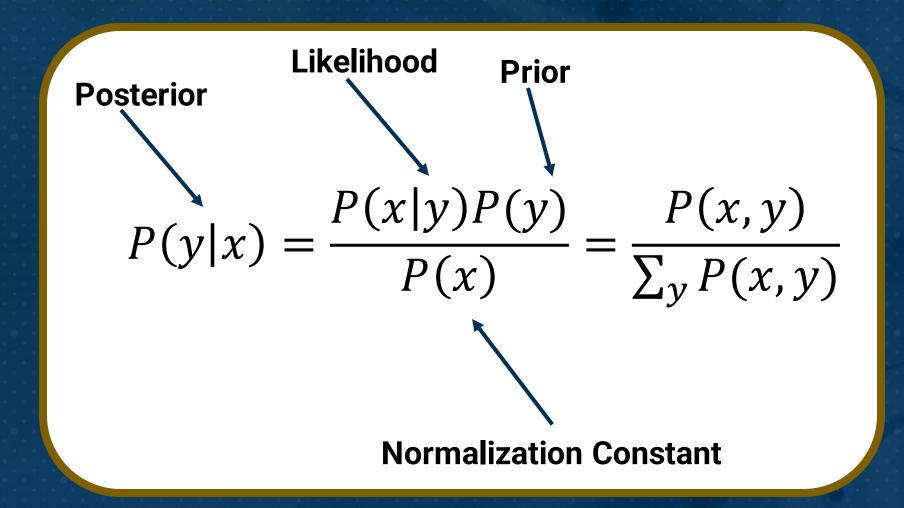
# **Learning Objectives**

In this lesson, you will learn a generative classifier

- Naïve Bayes
- Sentiment Analysis (positive, negative)
- Binary Classification



# Let's Start with the Math Concept: Bayes Decision Rule



- x is a document encoded i.e., by BoW
- y is the label of the document i.e., document contains a positive or negative message



#### **Bayes Decision Rule**

- Learning: prior : p(y), class conditional distribution: p(x|y)
- The poster probability of a test point:

$$q_i(x) := P(y = i|x) = \frac{P(x|y)P(y)}{P(x)}$$

- Bayes decision rule:
  - If  $q_i(x) > q_j(x)$ , then y = i, otherwise y = j
- Alternatively:
  - If ratio  $l(x) = \frac{P(x|y=i)}{P(x|Y=j)} > \frac{P(y=j)}{P(y=i)}$ , then y = i, otherwise y = j
  - Or look at the log-likelihood ratio  $h(x) = -ln \frac{q_i(x)}{q_i(x)}$



## What do People do in Practice?

- Generative models
  - Model prior and likelihood explicitly
  - "Generative" means able to generate synthetic data points
  - Examples: Naïve Bayes, Hidden Markov Models
- Discriminative models
  - Directly estimate the posterior probabilities
  - No need to model underlying prior and likelihood distributions
  - Examples: Logistic Regression, SVM, Neural Networks



#### Generative Model: Naïve Bayes

Use Bayes decision rule for classification

$$P(y|x) = \frac{P(x|y) P(y)}{P(x)}$$

• But assume p(x|y=1) is fully factorized

Dimensions (unique words) are independent.

$$p(x|y = 1) = \prod_{i=1}^{d} p(x_i|y = 1)$$

 Or the variables corresponding to each dimension of the data are independent given the label



# "Naïve" Conditional Independence Assumption

$$P(y|x) = \frac{P(x|y)P(y)}{P(x)} = \frac{P(x,y)}{P(x)}$$

$$P(x|y_{label=1})P(y_{label=1}) = P(x, y_{label=1})$$

$$= P(x_1|y_{label=1})P(x_2|y_{label=1}) ... P(x_d|y_{label=1})P(y_{label=1})$$

$$= P(y_{label=1}) \prod_{i=1}^{d} P(x_i|y_{label=1})$$



#### **Example: Conditional Independence**

$$P(y|x) = \frac{P(x|y)P(y)}{P(x)} = \frac{P(x,y)}{P(x)}$$

Vocabulary V = [nice, give, us, this, is, ssn, information, job, a]

```
P(document|y = positive)P(y = positive)
= P(x = nice|y = positive)P(x = give|y = positive) ... P(y = positive)
```

$$P(document|y = negative)P(y = negative)$$
  
=  $P(x = nice|y = negative)P(x = give|y = negative) \dots P(y = negative)$ 



## How to Represent the Likelihood?

$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

Vocabulary V = [nice, give, us, this, is, ssn, information, job, a]

$$P(document|y = positive)P(y = positive)$$
  
=  $P(x = nice|y = positive)P(x = give|y = positive) ... P(y = positive)$ 

$$P(x = give | y = positive)$$

A common distribution in NLP for Naïve Bayes is Multinomial Distribution

$$P(x = nice | y = positive) = \frac{\text{count word } nice \text{ in all doucments with positive labels}}{\text{count all words with positive labels}}$$



#### Example

VocabularyV = [nice, give, us, this, is, ssn, information, job, a]

$$P(document|y = positive)P(y = positive)$$
  
=  $P(x = nice|y = positive)P(x = give|y = positive) ... P(y = positive)$ 

$$P(x = nice | y = positive) = \frac{\text{count word } nice \text{ in all doucments with positive labels}}{\text{count all words with positive labels}}$$

$$P(y = positive) = \frac{count # positive documents}{count # all documents}$$



## Advantages and Disadvantages of Naïve Bayes

- Advantages:
  - Simple and easy to implement
  - No training needed
  - Good results in general
- Disadvantages:
  - The position of the words in the document does not matter (BoW approach)
  - Conditional independence



# Summary

- Naïve Bayes
- A generative model

