

# Chapter 3 Generative Models

Siheng Zhang  
zhangsiheng@cvte.com

September 11, 2020

This part corresponds to **Chapter 24, 31 in UML, Chapter ? in PRML, Chapter ? in PGM**, and mainly answers the following questions:

- 
- 

## Contents

<b>1</b>	<b>Bayesian classifier</b>	<b>2</b>
<b>2</b>	<b>Density estimation</b>	<b>2</b>
2.1	Parametric methods . . . . .	2
2.2	Non-parametric methods . . . . .	2
<b>3</b>	<b>Bayesian Reasoning</b>	<b>2</b>
<b>4</b>	<b>PAC-Bayes</b>	<b>2</b>
<b>5</b>	<b>Generative models</b>	<b>2</b>
5.1	GMM (Gaussian mixture models) . . . . .	2
5.2	HMM (Hidden Markov models) . . . . .	2
5.3	<i>v.s.</i> discriminant models . . . . .	2
5.4	Naive Bayes to linear discriminant models . . . . .	2
<b>6</b>	<b>Exercises and solutions</b>	<b>2</b>

## 1 Bayesian classifier

Recall that the Bayes optimal classifier (*in Chapter 1, Ex6*) is:

$$h_{\text{Bayes}}(\mathbf{x}) = \arg \max_{y \in \{0,1\}} \mathcal{P}(Y = y | X = \mathbf{x})$$

To describe the posterior probability function we need  $2^d$  parameters, this implies that the number of examples we need grows exponentially with the number of features. To avoid this problem, we assume that given the label, the features are independent of each other, i.e.,

$$\mathcal{P}(X = \mathbf{x} | Y = y) = \prod_{i=1}^d \mathcal{P}(X_i = \mathbf{x}_i | Y = y)$$

Together with Bayes' rule, the Bayes optimal classifier can be simplified as:

$$h_{\text{Bayes}}(\mathbf{x}) = \arg \max_{y \in \{0,1\}} \mathcal{P}(Y = y) \prod_{i=1}^d \mathcal{P}(X_i = \mathbf{x}_i | Y = y) \quad (1)$$

Now the number of parameters we need to estimate is only  $2d + 1$ . When we also estimate the parameters using the maximum likelihood principle (see below), the resulting classifier is called the *Naive Bayes* classifier.

## 2 Density estimation

### 2.1 Parametric methods

### 2.2 Non-parametric methods

## 3 Bayesian Reasoning

## 4 PAC-Bayes

## 5 Generative models

### 5.1 GMM (Gaussian mixture models)

### 5.2 HMM (Hidden Markov models)

### 5.3 *v.s.* discriminant models

### 5.4 Naive Bayes to linear discriminant models

The usual assumption in Naive Bayes classifier is that each conditional probability  $\mathcal{P}(X_i = \mathbf{x}_i | Y = y)$  is a Gaussian distribution. Consider the binary classification task, denote the

## 6 Exercises and solutions

*Chapter 4. Linear models, perceptron, MLP, deep learning, Generalization bounds on deep learning.*