ECE 595: Machine Learning I

Lecture 9.2: Bayesian Decision - Basic Principle

Spring 2020

Stanley Chan

School of Electrical and Computer Engineering
Purdue University

Outline

Generative Approaches

- Lecture 9 Bayesian Decision Rules
- Lecture 10 Evaluating Performance
- Lecture 11 Bayesian Parameter Estimation
- Lecture 12 Bayesian Prior
- Lecture 13 Connecting Bayesian and Linear Regression

Today's Lecture

- Review of High-Dimensional Gaussian
  - Likelihood and prior
  - Gaussian PDF
- Basic Principle
  - Making the Bayesian decision
  - 1D Illustration
- The Three Cases
  - $\boldsymbol{\Sigma}_i = \sigma^2 \boldsymbol{I}$
  - $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$ (Next Lecture)
  - General $\boldsymbol{\Sigma}_i$ (Next Lecture)

Interaction between Likelihood and Prior

- According to **Bayes Theorem**, we have that

$$p_{Y|\boldsymbol{X}}(i|\boldsymbol{x}) = \frac{p_{\boldsymbol{X}|Y}(\boldsymbol{x}|i)p_Y(i)}{p_{\boldsymbol{X}}(\boldsymbol{x})}$$

- Posterior: **After** you have seen $\boldsymbol{x}$
- Likelihood: **Before** you see $\boldsymbol{x}$
- Prior: You subjective believe of class label

- You cannot just use $p_Y(i)$; Otherwise you are not using data
- You cannot just use $p_{\boldsymbol{X}|Y}(\boldsymbol{x}|i)$; Otherwise you cannot explain "Y given $\boldsymbol{X}$"

Making the Bayesian Decision

Which class is more likely?

$$
\begin{aligned}
i^* &= \underset{i}{\text{argmax}}\ p_{Y|\boldsymbol{X}}(i|\boldsymbol{x}) \\
&= \underset{i}{\text{argmax}}\ \frac{p_{\boldsymbol{X}|Y}(\boldsymbol{x}|i)p_Y(i)}{p_{\boldsymbol{X}}(\boldsymbol{x})} \\
&= \underset{i}{\text{argmax}}\ \log p_{\boldsymbol{X}\,|\,Y}(\boldsymbol{x}|i) + \log \pi_i - \log p_{\boldsymbol{X}}(\boldsymbol{x}) \\
&= \underset{i}{\text{argmax}}\ \log p_{\boldsymbol{X}\,|\,Y}(\boldsymbol{x}|i) + \log \pi_i - \log p_{\boldsymbol{X}}(\boldsymbol{x})
\end{aligned}
$$

remove

- Solution = the most likely class according to posterior
- This involves a likelihood which depends on the model you choose
- This involves a prior term which is subjective

Let us Plug-in Multi-dimensional Gaussian

Recall d-dimensional Gaussian.

$$p_{\mathbf{X}\,|\,Y}(\mathbf{x}\,|\,i) = \frac{1}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}_i|}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right\}.$$

Plug this into the discriminant function

$$\begin{aligned} i^* &= \underset{i}{\text{argmax}}\ \log p_{\mathbf{X}\,|\,Y}(\mathbf{x}\,|\,i) + \log \pi_i \\ &= \underset{i}{\text{argmax}}\ -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) - \cancel{\frac{d}{2}\log(2\pi)} - \frac{1}{2}\log|\boldsymbol{\Sigma}_i| + \log \pi_i \\ &= \underset{i}{\text{argmax}}\ \underbrace{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)}_{\text{depend on } \mathbf{x}} \underbrace{-\frac{1}{2}\log|\boldsymbol{\Sigma}_i| + \log \pi_i.}_{\text{does not depend on } \mathbf{x}} \end{aligned}$$

Special Case: 1D; Two classes

The decision rule is

$$i^* = \underset{i}{\operatorname{argmax}} \underbrace{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)}_{\text{depend on } \boldsymbol{x}} \underbrace{-\frac{1}{2}\log|\boldsymbol{\Sigma}_i| + \log \pi_i}_{\text{does not depend on } \boldsymbol{x}}.$$

Substitute $\boldsymbol{\Sigma}_i = \sigma^2$, and $\boldsymbol{\mu}_i = \mu_i$. Do two classes.

$$-\frac{(x-\mu_1)^2}{2\sigma^2} - \log\sigma + \log\pi_1 \gtrless_{\mathcal{C}_2}^{\mathcal{C}_1} -\frac{(x-\mu_2)^2}{2\sigma^2} - \log\sigma + \log\pi_2$$

$$-\frac{(x-\mu_1)^2}{2\sigma^2} - \cancel{\log\sigma} + \log\pi_1 \gtrless_{\mathcal{C}_2}^{\mathcal{C}_1} -\frac{(x-\mu_2)^2}{2\sigma^2} - \cancel{\log\sigma} + \log\pi_2$$

$$\vdots$$

$$x \gtrless_{\mathcal{C}_2}^{\mathcal{C}_1} \underbrace{\frac{\mu_1-\mu_2}{2} - \frac{\sigma^2}{\mu_1-\mu_2}\log\frac{\pi_1}{\pi_2}}_{\text{does not depend on } x}.$$

Connecting to Linear Discriminant Function

Recall: A hypothesis function is

$$h(\boldsymbol{x}) = \begin{cases} 1, & \text{if} \quad g(\boldsymbol{x}) > 0 \\ 0, & \text{if} \quad g(\boldsymbol{x}) < 0 \\ \text{either}, & \text{if} \quad g(\boldsymbol{x}) = 0 \end{cases}$$

If there are only two classes, then we can define

$$g(\boldsymbol{x}) = g_i(\boldsymbol{x}) - g_j(\boldsymbol{x}).$$

where the i-th discriminant function is

$$g_i(\boldsymbol{x}) = \log p_{\boldsymbol{X}|Y}(\boldsymbol{x}|i) + \log \pi_i.$$

- Class i if $g(\boldsymbol{x}) > 0 \iff g_i(\boldsymbol{x}) > g_j(\boldsymbol{x})$
- Class j if $g(\boldsymbol{x}) < 0 \iff g_i(\boldsymbol{x}) < g_j(\boldsymbol{x})$
- Either if $g(\boldsymbol{x}) = 0 \iff g_i(\boldsymbol{x}) = g_j(\boldsymbol{x})$