Chapter 1:

Chapter 2: Optimal Classification

- **Error of classifier.**: $\epsilon[\psi(x)] = P(\psi(X) \neq Y) = \underbrace{p(\psi(X)=1|Y=0)}_{\epsilon^0=\int_{\{x|\psi(x)=1\}} p(x|Y=0)dx} P(Y=0) + \underbrace{p(\psi(X)=0|Y=1)}_{\epsilon^1=\int_{\{x|\psi(x)=0\}} p(x|Y=1)dx} P(Y=1)$
- **Cond. error**: $\epsilon[\psi|X] = P(\psi(X) \neq Y|X=x) = P(\psi(X)=0, Y=1|X=x) + P(\psi(X)=1, Y=0|X=x) = I_{\{\psi(x)=0\}}\eta(x) + I_{\{\psi(x)=1\}}(1-\eta(x))$
- **Post.prob.func.**: $\eta(x) = E[Y|X=x] = P(Y=1|X=x)$
- **Sensitivity**: $1-\epsilon^1[\psi]$; **Specificity**: $1-\epsilon^0[\psi]$
- **Thm.** *Bayes classifier*:

$$\psi^*(x) = \arg\max_i P(Y=i|X=x) = \begin{cases} 1, & \eta(x) > \frac{1}{2} \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

- **Thm. Bayes Error**: $\epsilon^* = P(Y=0)\epsilon^0[\psi^*] + P(Y=1)\epsilon^1[\psi^*] = E[\min\{\eta(X), 1-\eta(x)\}] = \frac{1}{2} - \frac{1}{2}E[|2\eta(X)-1|]$
- **Bayes class.**: $\psi^*(x) = \begin{cases} 1 & \overbrace{D^*(x)}^{\text{opt. discriminant}} > \overbrace{k^*}^{\text{opt. threshold}} \\ & = P(Y=1)p(x|Y=1) > P(Y=0)p(x|Y=0) \\ 0, & \text{otherwise} \end{cases}$
- $D^*(x) = \ln\frac{p(x|Y=1)}{p(x|Y=0)}$; $k^* = \ln\frac{P(Y=0)}{P(Y=1)}$

Gaussian Prob.:

$$p(x|Y=i) = \frac{1}{\sqrt{(2\pi)^d\det(\Sigma_i)}}\exp[\frac{1}{2}(x-\mu)^T\Sigma_i^{-1}(x-\mu_i)]$$

- $D^*(x) = \frac{1}{2}(x-\mu_0)^T\Sigma_0^{-1}(x-\mu_0) - \frac{1}{2}(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1) + \frac{1}{2}\ln\frac{\det(\Sigma_0)}{\det(\Sigma_1)}$

Homo. Case: Let $\|x_0-x_1\|_\Sigma = \sqrt{(x_0-x_1)^T\Sigma^{-1}(x_0-x_1)}$

$$\psi_L^*(x) = \begin{cases} 1, & \|x-\mu_1\|^2_\Sigma < \|x-\mu_0\|^2_\Sigma + 2\ln\frac{P(Y=1)}{P(Y=0)} \\ & = a^Tx + b > 0 \\ 0, & \text{otherwise} \end{cases}$$

- $a = \Sigma^{-1}(\mu_1-\mu_0)$ / $b = (\mu_0-\mu_1)^T\Sigma^{-1}(\frac{}{2})$; $b = (\mu_0-\mu_1)^T\Sigma^{-1}(\frac{\mu_0+\mu_1}{2}) + \ln\frac{P(Y=1)}{P(Y=0)}$
- $\epsilon_L^* = c\Phi(\frac{k^*-\frac{1}{2}\delta^2}{\delta}) + (1-c)\Phi(\frac{-k^*-\frac{1}{2}\delta^2}{\delta}), \delta = \sqrt{(\mu_1-\mu_0)^T\Sigma^{-1}(\mu_1-\mu_0)}$

Heter. Case: $\psi_Q^*(x) = \begin{cases} 1, & x^TAx + b^Tx + c > 0, \\ 0, & \text{otherwise} \end{cases}$

- $A = \frac{1}{2}(\Sigma_0^{-1} - \Sigma_1^{-1}); b = \Sigma_1^{-1}\mu_1 - \Sigma_0^{-1}\mu_0; c = \frac{1}{2}(\mu_0^T\Sigma_0^{-1}\mu_0 - \mu_1^T\Sigma_1^{-1}\mu_1) + \frac{1}{2}\ln\frac{\det\Sigma_0}{\det\Sigma_1} + \ln\frac{P(Y=1)}{P(Y=0)}$

Chapter 3: Sample-Based Classification

Chapter 4:

ffff

Chapter 5:

- Histogram Classification:

$$W_{n,h}(x, X_i) = \begin{cases} \frac{1}{N_h(x)}, & X_i \in A_h(x) \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

- **Thm. Cover-Hart**: $\epsilon_{NN} = E[2\eta(X)(1-\eta(x))]$

Bayes Error

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Key points & Definitions

- The posterior probability function is needed to define the Bayes classifier.
- Bayes error is optimal error
- LDA is parameteric

1. What are the minimum and the maximal values the Bayes error can take on in binary classification? Explain what each case means.

sample answer

2. Why is the expected classification error $\mu = E[\text{error}_n]$ not a function of the training data.
3. What does it mean to say that an error estimator is optimistically biased?
4. Is a consistent classification rule always better than a non-consistent one and why?
5. If a classifier is overfitted, will its apparent error (i.e., the error on the training data) tend to be smaller, larger, or the same as the true error? Explain why.
6. Describe the basic difference between filter and wrapper feature selection.
7. What is the penalty term in an SVM and what is it used for?
8. How many points does the minimal nonlinearly-separable problem in 2 dimensions have? Give an example.

Math facts

- Bayes: $P(Y=0|X=x) = \frac{P(Y=0)P(x|Y=0)}{P(x)}$
- $\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$; $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$