HullRobert-HW-1.md

Robert (Quinn) Hull

HW 1, ECE 523

2/1/2021

1 Probability and Discriminant Classifiers

Part 1: Maximum and Posterior v Probability of Chance

PART I: Maximum Posterior vs Probability of Chance Show/explain that $P(\omega_{\text{max}}|\mathbf{x}) \geq \frac{1}{c}$ when we are using the Bayes decision rule, where c is the number of classes. Derive an expression for $p(\text{err})$. Let ω_{max} be the state of nature for which $P(\omega_{\text{max}}|\mathbf{x}) \geq P(\omega_i|\mathbf{x})$ for $i =, 1 \ldots, c$. Show that $p(\text{err}) \leq (c-1)/c$ when we use the Bayes rule to make a decision. Hint, use the results from the previous questions.

Answer:

①

FIND $P(err)$ FACT: $P(w_{max}|x) \geq \frac{1}{c}$

a) The FACT $P(w_{max}|x) \geq \frac{1}{c}$ states in plan english that for a set of classes $w_1, w_2, w_3, \dots, w_c$ where x is a matrix of variables representing a sample and c is the number of classes (w):

- the maximum likelihood, of having class w_j given a set of variables x is at least as large as $1/c$.

- This is most intuitive when we think of chance. For example if you have a fair coin, the probability of heads v. tails is $1/c$, $c = 2$, $1/2$. Even if the coin were unfair, the maximum probability of the unfair side would have to be at least $1/2$, say $2/3$, or $3/4$.

b) Prove $p(err) \leq (c-1)/c$

recall: $p(x) = \sum_{y \in y} P(X, Y = y)$ (marginal distribution)

substituting "error" for X and x for Y, and using integral notation

$$p(error) = \int_x P(error, x)\, dx$$

using the product rule:

$$p(error) = \int_x P(error|x)\, p(x) \cdot dx$$

$$= \int_x (1 - p(w_{max}|x))\, p(x)\, dx$$

$$= \int_x p(x)\, dx - \int_x p(w_{max}|x)\, dx$$

- given $p(error|x) + p(w_{max}|x) = 1$
- using properties of integrals
- where $p(w_{max}|x) \geq \frac{1}{c}$
- where $p(x) = 1$

$$\boxed{P(err) \leq 1 - \frac{1}{c}}$$

← proof Complete

Part 2: Bayes Decision Rule Classifier

PART II: Bayes Decision Rule Classifier Let the elements of a vector $\mathbf{x} = [x_1, \ldots, x_d]^\top$ be binary valued. Let $P(\omega_j)$ be the prior probability of the class ω_j ($j \in [c]$), and let

$$p_{ij} = P(x_i = 1|\omega_j)$$

with all elements in $\mathbf{x}$ being independent. If $P(\omega_1) = P(\omega_2) = \frac{1}{2}$, and $p_{i1} = p > \frac{1}{2}$ and $p_{i2} = 1 - p$, show that the minimum error decision rule is

$$\text{Choose } \omega_1 \text{ if } \sum_{i=1}^{d} x_i > \frac{d}{2}$$

Hint: Think back to ECE503 and types of random variables then start out with

$$\text{Choose } \omega_1 \text{ if } P(\omega_1)P(\mathbf{x}|\omega_1) > P(\omega_2)P(\mathbf{x}|\omega_2)$$

Answer

II $P(w_1) = P(w_2) = 1/2$

$p = P_{i1} (> \frac{1}{2})$ The probability that $x_i = 1$ given class $= 1$

$1 - p = P_{i2} (< \frac{1}{2})$ The probability that $x_i = 1$ given class $= 2$

Show that Decision rule is:

Choose w_1 if $\sum_{i=1}^{d} x_i > \frac{d}{2}$, $\Leftrightarrow$ choose w_1 if more than half of elements in X are 1

- $\cancel{p(w_1)} P(x|w_1) > \cancel{p(w_2)} P(X|w_2)$
- $p(x|w_1) > p(x|w_2)$
- $\prod_{i=1}^{d} p(x_i|w_1) > \prod_{i=1}^{d} p(x_i|w_2)$

recall x is independent

drawing from the principles of Bernoulli, because this is a binary variable

$$p^K (1-p)^{(d-K)} > (1-p)^K p^{(d-K)}$$ where $K = \#$ successes

$$p^K (1-p)^d (1-p)^{-K} > (1-p)^K p^d p^{-K}$$

$$(1-p)^d p^{-d} > (1-p)^K p^{-K} p^{-K} (1-p)^K$$

$$(1-p)^d p^{-d} > (1-p)^{2K} p^{-2K}$$

$$\left(\frac{1-p}{p}\right)^d > \left(\frac{1-p}{p}\right)^{2K} \qquad f = \left(\frac{1-p}{p}\right)$$

$$f^d > f^{2K} \Rightarrow f^{(d-2K)} > 1$$

$$\left(\frac{1-p}{p}\right)^{(d-2K)} > 1$$

recall $p > \frac{1}{2}$, $p-1 < \frac{1}{2}$ so $\left(\frac{1-p}{p}\right) < 1$

Proof by trial & error
- if condition satisfied, choose w_1
- if condition not satisfied, choose w_2

~~Proof by trial and error~~

$$\left(\frac{1-p}{p}\right)^{d-2k} > 1 \quad , \text{ where } \left[\frac{1-p}{p} < 1\right] = f$$

① Condition $k = 0$

$$(f<1)^{d} > 1$$

Result: because exponent is a positive #, then left hand side does (L.H.S) < Right hand side (R.H.S). Choose w_2

This condition is if there 0 elements = 1

② Condition $k = d$

$$(f<1)^{(-d)} > 1$$

Result: exponent is negative so L.H.S. > R.H.S. Choose w_1

This condition if all elements = 1

③ Condition $k = \frac{d}{2}$

$$(f<1)^{0} > 1$$

Result: exponent is 0 so L.H.S. < R.H.S Choose w_2

This condition if half of elements = 1

from condition 3 we know if $k = \frac{d}{2}+1$, then we must choose w_1. In plain words, if more than half of elements in $x = 1$ then choose w_1.

This is identical to:

Choose w_1 if $\sum_{i=1}^{d} x_i > \frac{d}{2}$

Part 3: The Ditzler Household Growing Up

PART III: The Ditzler Household Growing Up My parents have two kids now grown into adults. Obviously there is me, Greg. I was born on a Wednesday. What is the probability that I have a brother? You can assume that $\mathbb{P}(\text{boy}) = \mathbb{P}(\text{girl}) = \frac{1}{2}$.

Answer

B, B

	M	T	W	Th	F	S	Su
M			x				
T			x				
W	x	x	x	x	x	x	x
Th			x				
F			x				
S			x				
Su			x				

13/49

B, G

	M	T	W	Th	F	S	Su
M							
T							
W	x	x	x	x	x	x	x
Th							
F							
S							
Su							

7/49

G, B

	M	T	W	Th	F	S	Su
M			x				
T			x				
W			x				
Th			x				
F			x				
Sa			x				
Su			x				

7/49

G, G

	M	T	W	Th	F	S	Su
M							
T							
W							
Th							
F							
S							
Su							

0/49

$P(B) = 1/2 \quad P(G) = 1/2$

$P(BB) = 1/4 \quad P(BG) = 1/4$

$P(GB) = 1/4 \quad P(GG) = 1/4$

We want: $P(BB|B) = \dfrac{P(B|BB) \cdot P(BB)}{P(B)}$

So $P(BB|B) = \dfrac{\Sigma BB}{\Sigma all} = \dfrac{13}{13+14} = \dfrac{13}{27} = 48\%$

Part 4: Bayes Classifier

PART IV: Bayes classifier Let consider a Bayes classifier with $p(\mathbf{x}|\omega_i)$ distributed as a multivariate Gaussian with mean μ_i and covariance $\Sigma_i = \sigma^2 I$ (note they all share the same covariance). We choose the class that has the largest

$$g_i(\mathbf{x}) = \log(p(\mathbf{x}|\omega_i)P(\omega_i)) \propto \mathbf{w}_i^\mathsf{T}\mathbf{x} + w_{0i}$$

Find $\mathbf{w}_i$ and w_{0i}. Fact:

$$p(\mathbf{x}|\omega_i) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_i|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu_i)^\mathsf{T}\Sigma_i^{-1}(\mathbf{x}-\mu_i)\right\}$$

Answer?

IV Bayes Classifier

prove that

$$g_i(x) = \underbrace{\log(p(x|w_i)Pw_i))}\ \propto W_i^T + W_{0i}$$

$$g_i(x) = \log(p(x|w_i)Pw_i)$$

$$g_i(x) = \log(p(x|w_i)) + \log(Pw_i)$$

We know $p(x|w_i) = \frac{1}{2\pi^{d/2}|\Sigma_i|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x-u_i)^T\Sigma_i^{-1}(x-u_i)\right]$

$$g_i(x) = \log\left(\frac{|\Sigma_i|^{-1/2}}{(2\pi)^{d/2}} \cdot \exp\left[-\frac{1}{2}(x-u_i)^T\Sigma_i^{-1}(x-u_i)\right]\right) + \log(P(w_i))$$

Now break up log

$$= \underbrace{\cancel{-\frac{1}{2}\log|\Sigma_i|}}_{1} - \underbrace{\cancel{\frac{d}{2}\log(2\pi)}}_{1} + \underbrace{\frac{1}{2}(x-u_i)^T\Sigma_i^{-1}(x-u_i)}_{2} + \log(P(w_i))$$

Now - simplify

1. remove constants
2. $\Sigma_i = \sigma^2 I$ where σ^2 are constant w/in and between classes, so sub σ^2

$$= \frac{-1}{2\sigma^2}(x-u_i)^T(x-u_i) + \log(P(w_i))$$

Further simplify that:

$$g_i x \quad \underset{(2)}{\frac{-1}{2\sigma^2}} \left(\cancel{x^T x}^{(1)} - 2u_i^T x + u_i^T u_i \right) + \log(P(w_i))$$

Our task is to maximize $g_i x$ for a given x, so
(1) $x^T x$ is constant for all class w_j. As such it reduces to (1) (is subtracted)
(2) gets distribut

is the same for all classes

$$g_i x \ \underset{\max}{\arg} \left\{ \frac{u_i^T x}{\sigma^2} - \frac{u_i^T u_i}{2\sigma^2} + \log(P(w_i)) \right\}$$

quick check for any more constants that wouldn't affect the comparison. I think though σ^2 is technically identical for each we need to keep it in there because it scales the bias depending on $\log(P(w_i))$. multiplying out

I think:

$$g_i x \ \underset{\max}{\arg} \left\{ u_i^T x - \underbrace{\frac{u_i^T u_i}{2} + \sigma^2 \cdot \log(P(w_i))} \right\}$$

where $W_i^T = u_i^T$, $W_{0i} = \frac{u_i^T u_i}{2} + \sigma^2 \log(P(w_i))$