

CMPS142 Homework 2

Connor Pryor
Student ID 1372839

Christopher Hsiao
Student ID 1398305

1. Naive Bayes

Use maximum likelihood estimation (not the unbiased or Laplace estimates) for the distributions of the two features conditioned on the two classes in the figure below. Give the mean and variance of the gaussians you found for the GPA.

label	AP	GPA
H	yes	4.0
H	yes	3.7
H	no	2.5
N	no	3.8
N	yes	3.3
N	yes	3.0
N	no	3.0
N	no	2.7
N	no	2.2

Mean and Variance

Honor Students:

$$mean = \frac{4.0 + 3.7 + 2.5}{3}$$

$$= 3.4$$

$$variance = \frac{(4.0 - 3.4)^2 + (3.7 - 3.4)^2 + (2.5 - 3.4)^2}{3}$$

$$= 0.42$$

Normal Students:

$$mean = \frac{3.8 + 3.3 + 3.0 + 3.0 + 2.7 + 2.2}{6}$$

$$= 3.0$$

$$\begin{aligned} \text{variance} &= \frac{(3.8 - 3.0)^2 + (3.3 - 3.0)^2 + (3.0 - 3.0)^2 + (3.0 - 3.0)^2 + (2.7 - 3.0)^2 + (2.2 - 3.0)^2}{6} \\ &= 0.243 \end{aligned}$$

Maximum Likelihood Estimation

One dimensional Gaussians:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x - \mu)^2}{2\sigma^2}}$$

Probability of AP:

Given H:

$$P(AP|H) = \frac{P(AP \text{ and } H)}{P(H)} = \frac{\frac{2}{9}}{\frac{3}{9}} = \frac{2}{3}$$

Given N:

$$P(AP|N) = \frac{P(AP \text{ and } N)}{P(N)} = \frac{\frac{2}{9}}{\frac{6}{9}} = \frac{1}{3}$$

Probability of \neg AP:

Given H:

$$P(\neg AP|H) = \frac{P(\neg AP \text{ and } H)}{P(H)} = \frac{\frac{1}{9}}{\frac{3}{9}} = \frac{1}{3}$$

Given N:

$$P(\neg AP|N) = \frac{P(\neg AP \text{ and } N)}{P(N)} = \frac{\frac{4}{9}}{\frac{6}{9}} = \frac{2}{3}$$

Predict Honor or Normal given AP and GPA:

To predict if a student is a Honor or Normal student given AP and a Gaussian distribution of the GPA. The decision boundary for Honor and Normal will be the same. This means the probability to be a Honor student given they have taken AP classes and a certain GPA will be equal to the probability to be a Normal student given they have taken AP classes and have the same GPA. Otherwise known as:

$$P(H|GPA, AP) = P(N|GPA, AP)$$

By Bayes Rule:

$$P(H|GPA, AP) = \frac{P(GPA, AP|H)P(H)}{P(GPA, AP|H)P(H) + P(GPA, AP|N)P(N)}$$

$$P(N|GPA, AP) = \frac{P(GPA, AP|N)P(N)}{P(GPA, AP|H)P(H) + P(GPA, AP|N)P(N)}$$

Since GPA and AP are independent, this can be rewritten as:

$$P(H|GPA, AP) = \frac{P(GPA|H)P(AP|H)P(H)}{P(GPA|H)P(AP|H)P(H) + P(GPA|N)P(AP|N)P(N)}$$

$$P(N|GPA, AP) = \frac{P(GPA|N)P(AP|N)P(N)}{P(GPA|H)P(AP|H)P(H) + P(GPA|N)P(AP|N)P(N)}$$

Since $P(H|GPA, AP) = P(N|GPA, AP)$ then:

$$P(GPA|H)P(AP|H)P(H) = P(GPA|N)P(AP|N)P(N)$$

Moving around the equation and substituting in $P(H)$, $P(N)$, $P(AP|H)$, and $P(AP|N)$

2. Gaussian Discriminant Analysis

a.) Show that the set of points \mathbf{x} such that $p(Y = 0|x) = p(Y = 1|x)$ is described by $\mathbf{w} \cdot \mathbf{x} = b$ for some \mathbf{w} and b .

We can assume that the label probabilities for $Y = 0$ and $Y = 1$ are the same. Which is to say

$$p(Y = 0) = p(Y = 1) = \frac{1}{2} \quad (1)$$

Using Bayes Theorem, the conditional probabilities given \mathbf{x} are as follows:

$$p(x|Y = 0) = \frac{(Y = 0|x)p(x)}{p(Y = 0)} ; p(x|Y = 1) = \frac{(Y = 1|x)p(x)}{p(Y = 1)} \quad (2)$$

Thus, we can deduce from (1) and (2) that:

$$p(Y = 0) = \frac{(Y = 0|x)p(x)}{p(x|Y = 0)} = \frac{(Y = 1|x)p(x)}{p(x|Y = 1)} = p(Y = 1) \quad (3)$$

We can deduce from this that:

$$p(Y = 1|x) \cdot p(x|Y = 0) = p(Y = 0|x) \cdot p(x|Y = 1) \quad (4)$$

We are attempting to classify values of x such that $p(Y = 0|x) = p(Y = 1|x)$, so we'll assume that this statement is true. Since these probabilities are equal, we have

$$p(Y = 0|x) = 1 - p(Y = 1|x) \quad (5)$$

Let $P = p(Y = 0|x)$. Then from (5), it follows that $P = 1 - P = p(Y = 1|x)$.

To sum up thus far, we have:

$$P \cdot p(x|Y = 0) = P \cdot p(x|Y = 1) \quad (6)$$

Returning to our original models since, we can now see that

$$\frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2}((x - \mu_0)^T \Sigma^{-1}(x - \mu_0))\right) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2}((x - \mu_1)^T \Sigma^{-1}(x - \mu_1))\right) \quad (7)$$

Which we can clearly cancel out the like terms on either side, and we have the equation

$$\exp\left(-\frac{1}{2}((x - \mu_0)^T \Sigma^{-1}(x - \mu_0))\right) = \exp\left(-\frac{1}{2}((x - \mu_1)^T \Sigma^{-1}(x - \mu_1))\right) \quad (8)$$

We can eliminate the exp and $-\frac{1}{2}$ by taking the log of both sides of the equation, then cancelling out $-\frac{1}{2}$. Now, the equation we're solving becomes:

$$(x - \mu_0)^T \Sigma^{-1}(x - \mu_0) = (x - \mu_1)^T \Sigma^{-1}(x - \mu_1) \quad (9)$$

By distributing the T ,

$$(x^T - \mu_0^T) \Sigma^{-1}(x - \mu_0) = (x^T - \mu_1^T) \Sigma^{-1}(x - \mu_1) \quad (10)$$

Now, we can multiply $\Sigma^{-1}(x - \mu_0)$ by $(x^T - \mu_0^T)$

$$x^T \Sigma^{-1}(x - \mu_0) - \mu_0^T \Sigma^{-1}(x - \mu_0) = x^T \Sigma^{-1}(x - \mu_1) - \mu_1^T \Sigma^{-1}(x - \mu_1) \quad (11)$$

Distributing all terms further, we arrive at

$$x^T \Sigma^{-1}x - x^T \Sigma^{-1}\mu_0 - \mu_0^T \Sigma^{-1}x + \mu_0^T \Sigma^{-1}\mu_0 = x^T \Sigma^{-1}x - x^T \Sigma^{-1}\mu_1 - \mu_1^T \Sigma^{-1}x + \mu_1^T \Sigma^{-1}\mu_1 \quad (12)$$

To reduce from here, we use the property of taking the transpose of a product (equation 18, below) to find that:

$$-2\mu_0^T \Sigma^{-1}x + \mu_0^T \Sigma^{-1}\mu_0 = -2\mu_1^T \Sigma^{-1}x + \mu_1^T \Sigma^{-1}\mu_1 \quad (13)$$

We group $-2\mu_0^T \Sigma^{-1}x$ and $-2\mu_1^T \Sigma^{-1}x$ and reduce to get

$$2(\mu_0^T \Sigma^{-1} - \mu_1^T \Sigma^{-1}) \cdot x = \mu_1^T \Sigma^{-1}\mu_1 - \mu_0^T \Sigma^{-1}\mu_0 \quad (14)$$

And by simply dividing the equation by 2, we end up with

$$(\mu_0^T \Sigma^{-1} - \mu_1^T \Sigma^{-1}) \cdot x = \frac{1}{2}(\mu_1^T \Sigma^{-1}\mu_1 - \mu_0^T \Sigma^{-1}\mu_0) \quad (15)$$

Observe that $\mu_0^T \Sigma^{-1} - \mu_1^T \Sigma^{-1}$ is a vector, as μ_0 and μ_1 are vectors. We will call this vector w .

Observe further that $\mu_1^T \Sigma^{-1}\mu_1 - \mu_0^T \Sigma^{-1}\mu_0$ is a scalar, because the product of a vector and its transpose yields a scalar. We will call this scalar b .

We now have the equation

$$w \cdot x = b \quad (16)$$

Recall that the original assumption made was that $p(Y = 0|x) = p(Y = 1|x)$. As such, this assumption implies that there exists some vector w and some scalar b such that $w \cdot x = b$ lies on the hyperplane.

Property of Taking the Transpose of a Product

Take note of the following property of taking the transpose of a product, that allows us to say:

$$(x^T \Sigma^{-1} \mu_i)^T = \mu_i^T \Sigma^{-1} x \quad (17)$$

Since Σ^{-1} is symmetric. Since we know that the product on the left side of (9) itself is a scalar, we can then also drop the transpose, where the equation then becomes:

$$x^T \Sigma^{-1} \mu_i = \mu_i^T \Sigma^{-1} x \quad (18)$$

b.) If $p(Y = 0) \neq p(Y = 1)$, then the reduction performed in part (*a.*) would not hold, as it could only be performed when we made the assumption that $p(Y = 0) = p(Y = 1)$. Thus, the decision boundary would no longer be a hyperplane.