Melissa LaHoud

Grand Canyon University

DSC 540: Machine Learning

Dr. Aiman Darwiche

4/7/2021

Part 1

In the case of normally distributed classes, discriminant functions are linear (straight lines, planes, and hyperplanes for two-, three-, and n-dimensional feature vectors, respectively) when the covariances matrices of corresponding classes are equal. Confirm this by deriving discriminant functions for a binary classification problem.

Given:

$$P(\mathbf{x}|y_q) = \frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}|^{1/2}} \exp\left(-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_q)^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_q)\right); q = 1, 2$$

Prove that linear discriminant functions

$$g_q(\mathbf{x}) = \boldsymbol{\mu}_q^T\mathbf{\Sigma}^{-1}\mathbf{x} - \tfrac{1}{2}\boldsymbol{\mu}_q^T\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_q + \ln P(y_q); q = 1, 2$$

And decision boundary $g(\mathbf{x}) = g_1(\mathbf{x}) - g_2(\mathbf{x}) = 0$ is given by

$$g(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0 = 0$$

$$\mathbf{w}^T\mathbf{x} + w_0 = (\boldsymbol{\mu}_1^T - \boldsymbol{\mu}_2^T)\mathbf{\Sigma}^{-1}\mathbf{x} - \tfrac{1}{2}(\boldsymbol{\mu}_1^T\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2^T\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_2) + \ln\frac{P(y_1)}{P(y_2)}$$

*(**Hint**: Use equations 3.61–3.62 in the textbook)*

Equation 3.61 in the textbook is below:

$$g_q(x) = lnP(y_q|x) or \ln p(x|y_q)P(y_q)$$

$$= ln\, p(x|y_q) + ln\, P(y_q); q = 1,2,..,M$$

From Bayes theorem, we have

$$P(y_q|x) = \frac{P(x|y_q)P(y_q)}{(2\pi)^{\frac{n}{2}}|\sum|^{\frac{1}{2}} P(x|y_q)P(y_q)}$$

$$= \frac{\exp\,(-\frac{1}{2}(x-\mu_q)^T\Sigma^{-1}(x-\mu_q)\; P(y_q)]}{(2\pi)^{\frac{n}{2}}|\sum|^{\frac{1}{2}} P(x|y_q)P(y_q)}$$

The denominator is the same in the above expression and decision boundary g(**x**) = g_1(**x**) – g_2(**x**) = 0, so we get the following:

$$\ln\left(P(y_q|x)\right) = -\frac{(x-\mu_q)^T\Sigma^{-1}(x-\mu_q)}{2} + \ln\,(P(y_q))$$

$$= -\frac{x^T\Sigma^{-1}x - 2x^T\Sigma^{-1}\mu_q + \mu_q^T\Sigma^{-1}\mu_q}{2} + \ln\,(P(y_q))$$

$x^T\Sigma^{-1}x$ is a constant for both classes so we can drop it from the expression:

$$g_q(x) = \ln\left(P(y_q|x)\right) = \mu_q^T\Sigma^{-1}x - \left(\frac{1}{2}\right)\mu_q^T\Sigma^{-1}\mu_q + \ln\left(P(y_q)\right) = g_2(x) = \ln(P(y_2|x))$$

$$= \mu_2^T\Sigma^{-1}x - \left(\frac{1}{2}\right)\mu_2^T\Sigma^{-1}\mu_2 + \ln(P(y_2))$$

After simplification we get the desired decision boundary.

$$g(x) = g_1(x) - g_2(x)$$

$$= \mu_1^T\Sigma^{-1}x - \left(\frac{1}{2}\right)\mu_1^T\Sigma^{-1}\mu_1 + \ln(P(y_1)) - \mu_2^T\Sigma^{-1}x - \left(\frac{1}{2}\right)\mu_2^T\Sigma^{-1}\mu_2 + \ln(P(y_2))$$

Regrouping the terms.

$$g(x) = (\mu_1^T - \mu_2^T)\Sigma^{-1}x - \frac{1}{2}(\mu_1^T\Sigma^{-1}\mu_1 - \mu_2^T\Sigma^{-1}\mu_2) + ln\frac{P(y_1)}{P(y_2)}$$

Part 2

Perform two iterations of the gradient algorithm to find the minima of $E(\boldsymbol{w}) = 2w_1^2 + 2w_1w_2 + 5w_2^2$.

The starting point is $\boldsymbol{w} = [\mathbf{2} \quad \mathbf{-2}]^{\boldsymbol{T}}$

Draw the contours and show your learning path graphically.

$$E(\boldsymbol{w}) = 2w_1^2 + 2w_1w_2 + 5w_2^2$$

$$\frac{\partial E(w)}{\partial w_1} = 4w_1 + 2w_2$$

$$\frac{\partial E(w)}{\partial w_2} = 2w_1 + 10w_2$$

Therefore, minimum of E(w) is

$$\nabla E(w) = (4w_1 + 2w_2)i + (2w_1 + 10w_2)j = 0$$

Let's start with the starting point $\boldsymbol{w} = [\mathbf{2} \quad \mathbf{-2}]^{\boldsymbol{T}}$ and alpha = 0.10

$$w_1 = w_1 - \propto \frac{\partial E(w)}{\partial w_1}$$

$$w_2 = w_2 - \propto \frac{\partial E(w)}{\partial w_2}$$

Iteration 1:

$$w_1 = 2 - (0.10)[4(2) + 2(-2)] = 1.6$$

$$w_2 = -2 - (0.10)[2(2) + 10(-2)] = -0.4$$

Iteration 2:

$$w_1 = 1.6 - (0.10)[4(1.6) + 2(-0.4)] = 1.04$$

$$w_2 = -0.4 - (0.10)[2(1.6) + 10(-0.4)] = -0.32$$

Lets Plot:

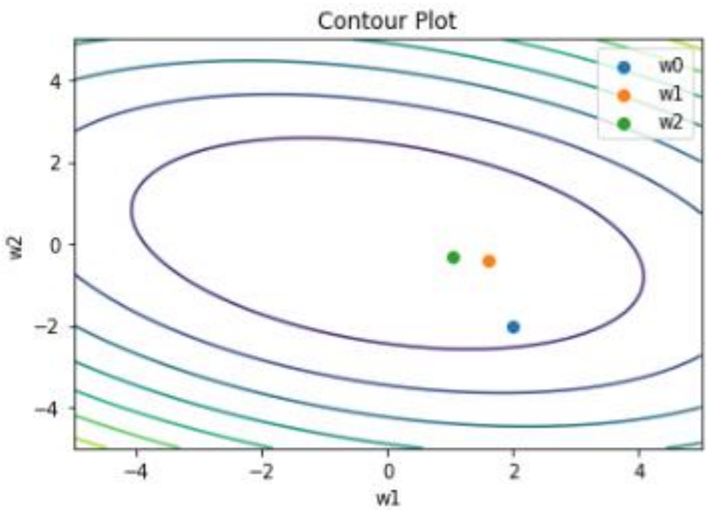

Figure 1: Contour Plot

As we can see from the Figure 1, the gradient goes to (0,0) which would be the minima of E(w).

Part 3

Show that logistic regression is a nonlinear regression problem. Is it possible to treat logistic discrimination in terms of an equivalent linear regression problem? Justify your answer.

According to Gopal, M. (2019), there is a way to adopt logistic regression models to classification problems which is a binary outcome. Classification problems usually are numerically coded such as 0 for 'yes' and 1 for 'no' and would not be correctly represented by linear regression. Probabilities are between 0 and 1 which is continuous so therefore it is a regression problem. "Odds is the ratio of probability of a particular outcome occurring to the probability of it not occurring. The odds is a number between 0 and infinity." (Global, M., 2019) The equation for a classification problem is below:

$$\log(odds) = log\frac{P(Class\ 1|x)}{1 - P(Class\ 1|x)} = w_0 + w_1x_1 + \cdots + w_nx_n$$

$w_0, w_1, \ldots, w_n$ are (n+1) parameters of linear logit function.

The result of the probability for P(Class 1|x) from the equation above is

$$P(Class\ 1|x) = \frac{1}{1 + \exp\ -(w_0 + w_1x_1 + \cdots + w_nx_n)}$$

$$P(Class\ 2|x) = 1 - P(Class\ 1|x)$$

The logistic function is always between 0 and 1 and we can use maximum likelihood to estimate parameters of logistic regression model.

References

Gopal, M. (2019). Applied Machine Learing. New Delhi, India: McGraw-Hill Education.