3.1 Pre-class Work

1. Barber Textbook Exercises

Exercise 17.1.

1. *Give an example of a two-dimensional dataset for which the data are linearly separable, but not linearly independent.*
2. *Can you find a dataset which is linearly independent but not linearly separable?*

17.1 Answers

1. {([1,1],0) , ([2,2],1)} → Can be seperated with a single line, however, [2,2] is a linear combination of [1,1]
2. Not sure.

Exercise 17.2. *Show that for both Ordinary ~~and Orthogonal~~ Least Squares regression fits to data $(x^n, y^n), n = 1, \ldots, N)$ the fitted lines go through the point $\sum_{n=1}^{N}(x^n, y^n)/N$.*

$\sum_{n=1}^{N}(x^n, y^n)/N = \bar{y}$

With the regression line, this is estimated by

$\bar{y} = \frac{1}{N}\sum_{n=1}^{N}\hat{y}_n + \epsilon_n$

$\hat{y}_n = \hat{\beta}_0 + \hat{\beta}_1 x_n$

Plugging in

$\bar{y} = \frac{1}{N}\sum_{n=1}^{N}\hat{\beta}_0 + \hat{\beta}_1 x_n + \hat{\epsilon}_n$

Error terms add up to zero

We get:

$\bar{y} = \hat{\beta}_0 + \frac{\hat{\beta}_1 \cdot \sum_{n=1}^{N} x_n}{N}$

$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$

This shows that it passes through the mean

Exercise 17.6. *The logistic sigmoid function is defined as $\sigma(x) = e^x/(1+e^x)$. What is the inverse function, $\sigma^{-1}(x)$?*

$\sigma(x) = e^x(1 + e^x)^{-1}$

A function g is the inverse of function f if for y=f(x), x=g(y) y = $e^x(1+e^x)^{-1}$

Replace x with y, and y with x

$x = \frac{e^y}{1+e^y}$

$e^y = x + xe^y$

$e^y - xe^y = x$

$e^y = \frac{x}{(1-x)}$

$y = ln\left(\frac{x}{(1-x)}\right)$

2. Murphy Textbook Exercises

Exercise 8.3 Gradient and Hessian of log-likelihood for logistic regression

a. Let $\sigma(a) = \frac{1}{1+e^{-a}}$ be the sigmoid function. Show that

$$\frac{d\sigma(a)}{da} = \sigma(a)(1-\sigma(a)) \tag{8.124}$$

b. Using the previous result and the chain rule of calculus, derive an expression for the gradient of the log likelihood (Equation 8.5).

Answers

a. $\sigma(a) = \frac{1}{1+e^{-a}} = (1+e^{-a})^{-1}$

Apply chain rule:

$= -(1+e^{-a})^{-2} \quad \cdot \frac{\partial}{\partial a}(1+e^{-a})$

$= -(1+e^{-a})^{-2} \quad \cdot -e^{-a}$

$= \frac{e^{-a}}{(1+e^{-a})^2} = \frac{1}{1+e^{-a}} \frac{e^{-a}}{1+e^{-a}}$

= $\sigma(a)(1-\sigma(a))$

b.

$g(w) = \frac{\partial}{\partial a} LL$

$$= \sum_{i=1}^{N} \frac{\partial}{\partial w}[y_i log \mu_i + (1-y_i)log(1-\mu_i)]$$