CS446: Machine Learning, Fall 2017, Homework 1

Name: Yiming Gao (yimingg2)

Worked individually

Relation Between Logistic Regression and Naive Bayes

Problem 1

Solution:
From Bayes rule, we know

$$P(y=1|\boldsymbol{x}) = \frac{P(y=1)P(\boldsymbol{x}|y=1)}{P(y=1)P(\boldsymbol{x}|y=1)+P(y=0)P(\boldsymbol{x}|y=0)}$$

Problem 2

Solution: https://www.cs.cmu.edu/ tom/mlbook/NBayesLogReg.pdf
Dividing both the numerator and denominator by the numerator yields:

$$P(y=1|\boldsymbol{x}) = \frac{1}{1+\frac{P(y=0)P(\boldsymbol{x}|y=0)}{P(y=1)P(\boldsymbol{x}|y=1)}}$$

or equivalently

$$\begin{aligned}
P(y=1|\boldsymbol{x}) &= \frac{1}{1+exp(ln\frac{P(y=0)P(\boldsymbol{x}|y=0)}{P(y=1)P(\boldsymbol{x}|y=1)})} \\
&= \frac{1}{1+exp(ln\frac{1-\pi}{\pi}+ln\frac{P(\boldsymbol{x}|y=0)}{P(\boldsymbol{x}|y=1)})}
\end{aligned}$$

is in the sigmoid function form $\sigma(a) = \frac{1}{1+exp^{-a}}$ where $a = -(ln\frac{1-\pi}{\pi}+ln\frac{P(\boldsymbol{x}|y=0)}{P(\boldsymbol{x}|y=1)})$.

Problem 3

Solution:
Given label $y=c$, each $x_i \in \boldsymbol{x}$ has a Gaussian distribution, i.e., $x_i \sim N(\mu_{ic}, \sigma_i^2)$.

$$
\begin{aligned}
P(\boldsymbol{x}|y=c) &= \prod_{i=1}^{d} P(x_i|y=c) \\
&= \prod_{i=1}^{d} \frac{1}{\sqrt{2\pi\sigma_i^2}} exp(-\frac{(x_i-\mu_{ic})^2}{2\sigma_i^2})
\end{aligned}
$$

Problem 4

Solution:

Substituting the result of part 3 to part 2, we have

$$
\begin{aligned}
P(y=1|\boldsymbol{x}) &= \frac{1}{1+exp(ln\frac{P(y=0)P(\boldsymbol{x}|y=0)}{P(y=1)P(\boldsymbol{x}|y=1)})} \\
&= \frac{1}{1+exp(ln\frac{1-\pi}{\pi}+ln\frac{P(\boldsymbol{x}|y=0)}{P(\boldsymbol{x}|y=1)})}
\end{aligned}
$$

First we consider the production term, which is in summation form after taking log:

$$
\begin{aligned}
ln\frac{P(\boldsymbol{x}|y=0)}{P(\boldsymbol{x}|y=1)} &= \sum_i ln\frac{P(x_i|y=0)}{P(x_i|y=1)} \\
&= \sum_i ln\frac{\frac{1}{\sqrt{2\pi\sigma_i^2}}exp(-\frac{(x_i-\mu_{i0})^2}{2\sigma_i^2})}{\frac{1}{\sqrt{2\pi\sigma_i^2}}exp(-\frac{(x_i-\mu_{i1})^2}{2\sigma_i^2})} \\
&= \sum_i \frac{(x_i-\mu_{i1})^2-(x_i-\mu_{i0})^2}{2\sigma_i^2} \\
&= \sum_i \frac{(x_i^2-2x_i\mu_{i1}+\mu_{i1}^2)-(x_i^2-2x_i\mu_{i0}+\mu_{i0}^2)}{2\sigma_i^2} \\
&= \sum_i \frac{2x_i(\mu_{i0}-\mu_{i1})+\mu_{i1}^2-\mu_{i0}^2}{2\sigma_i^2} \\
&= \sum_i (\frac{\mu_{i0}-\mu_{i1}}{\sigma_i^2}x_i+\frac{\mu_{i1}^2-\mu_{i0}^2}{2\sigma_i^2})
\end{aligned}
$$

Note this expression is a linear weighted sum of the x_i's. We have

$$
P(y=1|\boldsymbol{x}) = \frac{1}{1+exp(ln\frac{\pi}{1-\pi}+\sum_i(\frac{\mu_{i0}-\mu_{i1}}{\sigma_i^2}x_i+\frac{\mu_{i1}^2-\mu_{i0}^2}{2\sigma_i^2}))}
$$

Or equivalently,

$$
P(y=1|\boldsymbol{x}) = \frac{1}{1+e^{-(w_0+\sum_{i=1}^{d} w_i x_i)}} = \frac{1}{1+e^{-\boldsymbol{w}^T\boldsymbol{x}}}
$$

where the weights $w_1, w_2, ..., w_d$ are given by

$$w_i = \frac{\mu_{i1} - \mu_{i0}}{\sigma_i^2}$$

and where

$$w_0 = -ln\frac{1-\pi}{\pi} + \sum_i \frac{\mu_{i0}^2 - \mu_{i1}^2}{2\sigma_i^2}$$

.