

Machine Learning HW3

Wenhan Yang

Due: March 28 2021

Problem 1: Logistic Regression

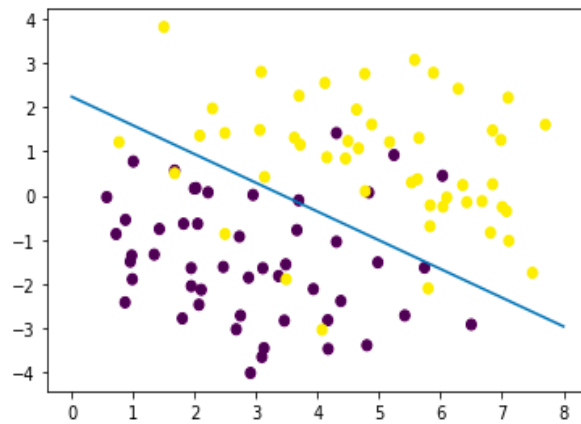
1.

Model coefficients: $w_0/\text{intercept} = [-2.62004691]$, $w_1 = [0.76026934]$, $w_2 = [1.17191257]$

2.

Coefficients of linear decision boundary: $\text{intercept} = [2.23570168]$, $\text{slope} = [-0.64874237]$

Training data and decision boundary:



Problem 2: Linear Discriminative Analysis

1.

$$\begin{cases} P(y=1|x) = \frac{P(x|y=1) * P(y=1)}{P(x|y=1) * P(y=1) + P(x|y=0) * P(y=0)} \\ P(x|y=0) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp(-\frac{1}{2}(x - \mu_0)^T \Sigma^{-1} (x - \mu_0)) \\ P(x|y=1) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp(-\frac{1}{2}(x - \mu_1)^T \Sigma^{-1} (x - \mu_1)) \end{cases}$$

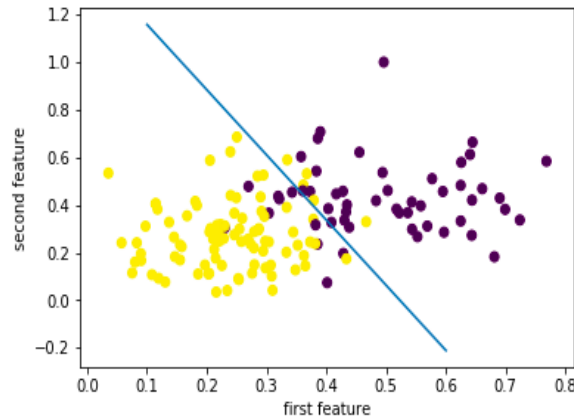
Given $\phi, \mu_0, \mu_1, \Sigma$ learned, the posterior distribution can be written as:

$$\begin{aligned} P(y=1|x; \mu_0, \mu_1, \Sigma) &= \frac{P(x|y=1) * P(y=1)}{P(x|y=1) * P(y=1) + P(x|y=0) * P(y=0)} \\ &= \frac{\frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp(-\frac{1}{2}(x - \mu_1)^T \Sigma^{-1} (x - \mu_1)) * \phi}{\frac{\exp(-\frac{1}{2}(x - \mu_1)^T \Sigma^{-1} (x - \mu_1))}{(2\pi)^{d/2} |\Sigma|^{1/2}} * \phi + \frac{\exp(-\frac{1}{2}(x - \mu_0)^T \Sigma^{-1} (x - \mu_0))}{(2\pi)^{d/2} |\Sigma|^{1/2}} * (1 - \phi)} \\ &= \frac{\exp(-\frac{1}{2}(x - \mu_1)^T \Sigma^{-1} (x - \mu_1)) * \phi}{\exp(-\frac{1}{2}(x - \mu_1)^T \Sigma^{-1} (x - \mu_1)) * \phi + \exp(-\frac{1}{2}(x - \mu_0)^T \Sigma^{-1} (x - \mu_0)) * (1 - \phi)} \\ &= \frac{1}{1 + \exp\left(-\frac{1}{2}(x^T - \mu_0^T) \Sigma^{-1} (x - \mu_0) + \frac{1}{2}(x^T - \mu_1^T) \Sigma^{-1} (x - \mu_1)\right)} * \frac{1 - \phi}{\phi} \\ &= \frac{1}{1 + \exp\left(-\frac{1}{2}\left(-2\mu_0^T \Sigma^{-1} x + 2\mu_1^T \Sigma^{-1} x + \mu_0^T \Sigma^{-1} \mu_0 - \mu_1^T \Sigma^{-1} \mu_1\right) + \ln\left(\frac{1 - \phi}{\phi}\right)\right)} \\ &= \frac{1}{1 + \exp\left(\mu_0^T \Sigma^{-1} x - \mu_1^T \Sigma^{-1} x - \frac{1}{2}\mu_0^T \Sigma^{-1} \mu_0 + \frac{1}{2}\mu_1^T \Sigma^{-1} \mu_1 + \ln\left(\frac{1 - \phi}{\phi}\right)\right)} \\ &= \frac{1}{1 + \exp\left((\mu_0 - \mu_1)^T \Sigma^{-1} x + \frac{1}{2}(\mu_1^T \Sigma^{-1} \mu_1 - \mu_0^T \Sigma^{-1} \mu_0) + \ln\left(\frac{1 - \phi}{\phi}\right)\right)} \end{aligned}$$

$$\text{where } \begin{cases} \theta_0 = \frac{1}{2}(\mu_0^T \Sigma^{-1} \mu_0 - \mu_1^T \Sigma^{-1} \mu_1) + \ln\left(\frac{\phi}{1 - \phi}\right) \\ \theta^T = (\mu_0 - \mu_1)^T \Sigma^{-1} \end{cases}$$

2.

the accuracy score is 0.8951048951048951.



coefficients of linear decision boundary: intercept = 1.4311620254386697, slope = -2.7373690432520297

Problem 3: Naive Bayes

1.

Total # of words: 3048

Validation error, # = 61, % = 6.1000%.

2.

With threshold 1....

train:12.275 val:14.7 test:16.5 len(V)=65812

time: 2396.615859746933

With threshold 2....

train:4.45 val:6.1 test:7.3 len(V)=19298

time: 709.3737497329712

With threshold 3....

train:4.0 val:5.4 test:5.9 len(V)=12882

time: 474.42129373550415

With threshold 4....

train:3.975 val:5.5 test:5.3 len(V)=10362

time: 380.69529151916504

With threshold 5....

train:3.95 val:5.5 test:5.2 len(V)=8773

time: 322.3561816215515

With threshold 6....

train:4.1 val:5.6 test:5.4 len(V)=7803

time: 286.11717438697815

With threshold 7....

train:4.225 val:5.5 test:5.4 len(V)=7085

time: 259.6546537876129

With threshold 8....

train:4.2 val:5.1 test:5.5 len(V)=6493

time: 238.80849313735962

With threshold 9....

train:4.4 val:5.1 test:6.0 len(V)=5982

time: 218.68049216270447

With threshold 10....

train:4.55 val:5.0 test:6.1 len(V)=5603

time: 205.38423895835876

With threshold 11....

train:4.55 val:5.1 test:6.2 len(V)=5303

time: 193.67969751358032

With threshold 12....

train:4.625 val:5.0 test:6.2 len(V)=4988

time: 182.57376670837402

With threshold 13....

train:4.65 val:5.1 test:6.4 len(V)=4727

time: 172.415105342865

With threshold 14....

train:4.775 val:5.3 test:6.3 len(V)=4499

time: 164.35581350326538

With threshold 15....

train:4.725 val:5.5 test:6.4 len(V)=4310

time: 156.6315586566925

With threshold 16....

train:4.725 val:5.6 test:6.3 len(V)=4133

time: 150.6678183078766
With threshold 17....
train:4.8 val:5.5 test:6.4 len(V)=3984
time: 144.4842562675476
With threshold 18....
train:4.85 val:5.5 test:6.4 len(V)=3848
time: 139.93675303459167
With threshold 19....
train:5.025 val:5.6 test:6.6 len(V)=3724
time: 134.8518669605255
With threshold 20....
train:5.05 val:5.7 test:6.7 len(V)=3613
time: 131.36252880096436
With threshold 21....
train:5.125 val:5.7 test:6.8 len(V)=3521
time: 127.3626606464386
With threshold 22....
train:5.125 val:5.7 test:6.8 len(V)=3408
time: 123.13043785095215
With threshold 23....
train:5.175 val:5.7 test:6.9 len(V)=3308
time: 119.47244238853455
With threshold 24....
train:5.275 val:5.8 test:7.0 len(V)=3206
time: 115.69488406181335
With threshold 25....
train:5.525 val:6.0 test:7.2 len(V)=3137
time: 113.56705689430237
With threshold 26....
train:5.575 val:6.1 test:7.4 len(V)=3048
time: 109.89387154579163
With threshold 27....
train:5.6 val:6.1 test:7.6 len(V)=2968
time: 107.30016040802002
With threshold 28....
train:5.7 val:6.2 test:7.7 len(V)=2906
time: 104.79076480865479
With threshold 29....
train:5.725 val:6.3 test:7.8 len(V)=2840
time: 102.48775863647461
With threshold 30....
train:5.7 val:6.2 test:7.6 len(V)=2769
time: 99.66372466087341
With threshold 31....
train:5.7 val:6.2 test:7.5 len(V)=2712
time: 97.77890944480896
With threshold 32....
train:5.75 val:6.1 test:7.5 len(V)=2655
time: 95.4313383102417
With threshold 33....
train:5.725 val:6.1 test:7.6 len(V)=2607
time: 93.85421395301819
With threshold 34....
train:5.75 val:6.1 test:7.7 len(V)=2562
time: 91.9760468006134

Best performance at validated threshold 10 with test error rate 6.1.

