CSE 242 Homework 2

Sampson Mao, samao@ucsc.edu

November 7, 2020

1 Logistic regression

1.1 (a)

$$
\begin{aligned}
q = g(w^T x) &= \frac{e^{w^T x}}{1 + e^{w^T x}} \\
\frac{1}{q} = \frac{1 + e^{w^T x}}{e^{w^T x}} &= \frac{1}{e^{w^T x}} + 1 \\
\frac{1}{q} - 1 &= \frac{1}{e^{w^T x}} \\
\log\left(\frac{1-q}{q}\right) &= -\log e^{w^T x} \\
-\log\left(\frac{1-q}{q}\right) &= w^T x \\
\implies \log\left(\frac{q}{1-q}\right) &= w^T x
\end{aligned}
$$

1.2 (b)

If we view g in terms of the likelihood, then g is a function of the vector parameter $\boldsymbol{w}$. Below, I will use properties of vector derivatives to arrive at the result.

Let $\ell(\boldsymbol{w}; \boldsymbol{x}) = \log g(\boldsymbol{w}^T \boldsymbol{x})$

$$
\begin{aligned}
\frac{\partial \ell(\boldsymbol{w}; \boldsymbol{x})}{\partial \boldsymbol{w}} &= \frac{\partial}{\partial \boldsymbol{w}} \left(\boldsymbol{w}^T \boldsymbol{x} - \log(1 + e^{\boldsymbol{w}^T \boldsymbol{x}})\right) \\
&= \frac{\partial}{\partial \boldsymbol{w}} \boldsymbol{w}^T \boldsymbol{x} - \frac{\partial}{\partial \boldsymbol{w}} \log(1 + e^{\boldsymbol{w}^T \boldsymbol{x}}) \\
&= \boldsymbol{x} - \frac{e^{\boldsymbol{w}^T \boldsymbol{x}}}{1 + e^{\boldsymbol{w}^T \boldsymbol{x}}} \boldsymbol{x} && \left(\frac{\partial \boldsymbol{w}^T \boldsymbol{x}}{\partial \boldsymbol{w}} = \boldsymbol{x}\right) \\
&= \boldsymbol{x} \left(1 - \frac{e^{\boldsymbol{w}^T \boldsymbol{x}}}{1 + e^{\boldsymbol{w}^T \boldsymbol{x}}}\right) \\
&= \boldsymbol{x} \left(1 - g(\boldsymbol{w}^T \boldsymbol{x})\right)
\end{aligned}
$$

2 Naive Bayes

We would like to predict the quantity $P(Type|GPA, AP)$, where $Type$ can be honors H or normal N. Using Bayes rule,

$$P(Type|GPA, AP) = \frac{P(GPA, AP|Type)P(Type)}{P(GPA, AP)}$$

Then by conditional independence of GPA and AP (the Naive assumption),

$$P(Type|GPA, AP) = \frac{P(GPA|Type)P(AP|Type)P(Type)}{P(GPA, AP)}$$

From the problem we have that the conditional distribution of the GPA is a Gaussian distribution and the conditional distribution of taking AP classes is a Bernoulli distribution.

The likelihood of the conditional Gaussian distribution using the data is as follows. Here, n is the number of Types in the honors class when we are finding $P(H|GPA, AP)$ and the number in the normal class when finding $P(N|GPA, AP)$

$$\begin{aligned}
P(GPA|Type) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(GPA_i - \mu)^2}{2\sigma^2}\right\} \\
\log P(GPA|Type) &= \sum_{i=1}^{n} \left(-\frac{1}{2}\log 2\pi\sigma^2 - \frac{(GPA_i - \mu)^2}{2\sigma^2}\right)
\end{aligned}$$

Finding $\hat{\mu}_{MLE}$ involves finding the derivative with respect to μ in the squared term, which we can simplify first. Denote $\overline{GPA} = \frac{1}{n}\sum_{i=1}^{n} GPA_i$

$$\begin{aligned}
\sum_{i=1}^{n}(GPA_i - \mu)^2 &= \sum_{i=1}^{n}(GPA_i - \overline{GPA} + \overline{GPA} - \mu)^2 \\
&= \sum_{i=1}^{n}(GPA_i - \overline{GPA})^2 - 2\sum_{i=1}^{n}(GPA_i - \overline{GPA})(\overline{GPA} - \mu) + \sum_{i=1}^{n}(\overline{GPA} - \mu)^2 \\
&= \sum_{i=1}^{n}(GPA_i - \overline{GPA})^2 - 2(n \cdot \overline{GPA} - n \cdot \overline{GPA})(\overline{GPA} - \mu) + \sum_{i=1}^{n}(\overline{GPA} - \mu)^2 \\
&= \sum_{i=1}^{n}(GPA_i - \overline{GPA})^2 + n(\overline{GPA} - \mu)^2 \\
&\propto (\overline{GPA} - \mu)^2
\end{aligned}$$

Taking the derivative with respect to μ, we find that $\hat{\mu}_{MLE} = \overline{GPA}$.

For $\hat{\sigma^2}_{MLE}$, first simplify the log-likelihood to keep terms important to σ^2

$$\log P(GPA|H) = \sum_{i=1}^{n}\left(-\frac{1}{2}\log 2\pi\sigma^2 - \frac{(GPA_i-\mu)^2}{2\sigma^2}\right)$$
$$\propto -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(GPA_i-\mu)^2$$

$$\frac{\partial}{\partial\sigma^2}\log P(GPA|H) = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(GPA_i-\mu)^2 = 0$$
$$\frac{1}{(\sigma^2)^2}\sum_{i=1}^{n}(GPA_i-\mu)^2 = \frac{n}{\sigma^2}$$
$$\hat{\sigma^2}_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(GPA_i-\hat{\mu}_{MLE})^2$$

Meanwhile the AP probability has a Bernoulli distribution, so the pmf is

$$P(AP|Type) = p^{AP}(1-p)^{1-AP}$$

where $AP = 1$ is having taken the AP course and $AP = 0$ otherwise. $\hat{p}_{MLE}$ is the proportion of honors or non-honors students that have taken AP courses.

So for the honors case,

$$\hat{\mu}_{H,MLE} = \frac{4+3.7+2.5}{3} = 3.4\hat{\sigma^2}_{H,MLE} = \frac{(4-3.4)^2+(3.7-3.4)^2+(2.5-3.4)^2}{3} = 0.42\hat{p}_{H,MLE} = \frac{2}{3}$$

Likewise for the normal case,

$$\hat{\mu}_{N,MLE} = 3\hat{\sigma^2}_{N,MLE} = 0.243\hat{p}_{H,MLE} = \frac{2}{6}$$

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

$$= \frac{\frac{1}{\sqrt{2\pi\cdot 0.42}}\exp\left\{-\frac{(GPA-3.4)^2}{2\cdot 0.42}\right\}\left(\frac{2}{3}\right)^{AP}\left(\frac{1}{3}\right)^{1-AP}\left(\frac{3}{9}\right)}{\frac{1}{\sqrt{2\pi\cdot 0.42}}\exp\left\{-\frac{(GPA-3.4)^2}{2\cdot 0.42}\right\}\left(\frac{2}{3}\right)^{AP}\left(\frac{1}{3}\right)^{1-AP}\left(\frac{3}{9}\right)+\frac{1}{\sqrt{2\pi\cdot 0.243}}\exp\left\{-\frac{(GPA-3)^2}{2\cdot 0.243}\right\}\left(\frac{1}{3}\right)^{AP}\left(\frac{2}{3}\right)^{1-AP}\left(\frac{6}{9}\right)}$$

$$P(H|GPA, AP = 1) = \frac{\frac{1}{\sqrt{2\pi \cdot 0.42}} \exp\left\{-\frac{(GPA-3.4)^2}{2\cdot 0.42}\right\}\left(\frac{2}{3}\right)\left(\frac{3}{9}\right)}{\frac{1}{\sqrt{2\pi \cdot 0.42}} \exp\left\{-\frac{(GPA-3.4)^2}{2\cdot 0.42}\right\}\left(\frac{2}{3}\right)\left(\frac{3}{9}\right) + \frac{1}{\sqrt{2\pi \cdot 0.243}} \exp\left\{-\frac{(GPA-3)^2}{2\cdot 0.243}\right\}\left(\frac{1}{3}\right)\left(\frac{6}{9}\right)}$$

$$0.5 \geq \frac{0.1368 \exp\left\{-\frac{(GPA-3.4)^2}{0.84}\right\}}{0.1368 \exp\left\{-\frac{(GPA-3.4)^2}{0.84}\right\} + 0.1797 \exp\left\{-\frac{(GPA-3)^2}{0.486}\right\}}$$

$$2 \geq 1 + \frac{0.1797 \exp\left\{-\frac{(GPA-3)^2}{0.486}\right\}}{0.1368 \exp\left\{-\frac{(GPA-3.4)^2}{0.84}\right\}}$$

$$0 \geq \log\left(\frac{0.1797}{0.1368}\right) - \frac{(GPA-3)^2}{0.486} + \frac{(GPA-3.4)^2}{0.84}$$

$$GPA \geq 3.365 \text{ or } GPA \leq 1.537$$

$$P(H|GPA, AP = 0) = \frac{\frac{1}{\sqrt{2\pi \cdot 0.42}} \exp\left\{-\frac{(GPA-3.4)^2}{2\cdot 0.42}\right\}\left(\frac{1}{3}\right)\left(\frac{3}{9}\right)}{\frac{1}{\sqrt{2\pi \cdot 0.42}} \exp\left\{-\frac{(GPA-3.4)^2}{2\cdot 0.42}\right\}\left(\frac{1}{3}\right)\left(\frac{3}{9}\right) + \frac{1}{\sqrt{2\pi \cdot 0.243}} \exp\left\{-\frac{(GPA-3)^2}{2\cdot 0.243}\right\}\left(\frac{2}{3}\right)\left(\frac{6}{9}\right)}$$

$$0.5 \geq \frac{0.06840 \exp\left\{-\frac{(GPA-3.4)^2}{0.84}\right\}}{0.06840 \exp\left\{-\frac{(GPA-3.4)^2}{0.84}\right\} + 0.3596 \exp\left\{-\frac{(GPA-3)^2}{0.486}\right\}}$$

$$2 \geq 1 + \frac{0.3596 \exp\left\{-\frac{(GPA-3)^2}{0.486}\right\}}{0.06840 \exp\left\{-\frac{(GPA-3.4)^2}{0.84}\right\}}$$

$$0 \geq \log\left(\frac{0.3596}{0.06840}\right) - \frac{(GPA-3)^2}{0.486} + \frac{(GPA-3.4)^2}{0.84}$$

$$GPA \geq 4.011 \text{ or } GPA \leq 0.890$$

Using a threshold of 0.50,

If AP courses are taken, predict H if the GPA is between 0 and 1.533, and 3.365 to 4;

if AP courses are not taken, predict H if the GPA is between 0 and 0.890. (GPA is from 0 to 4 only)

The calculations were done with Wolfram alpha.

3 Nearest Neighbor

3.1 (a)

Let's call Y the predicted label and t the actual label. We would like to find the quantity

$$P(Y \neq t)$$

So either the predicted value $y = +$ and $t = -$ or $y = -$ and $t = +$

$$\begin{aligned} P(Y \neq t) &= P(Y = +, t = -) + P(Y = -, t = +) \\ &= P(Y = +)P(t = -) + P(Y = -)P(t = +) \end{aligned}$$

The predicted point comes from a majority vote (i.e. at least 2 of the points must be the same out of 3). Call the 3 nearest neighbors X_1, X_2, X_3

$$P(Y = y) = P(X_1 = y)P(X_2 = y)P(X_3 = y) + \binom{3}{1} P(X_1 = y)P(X_2 = y)P(X_3 = (1 - y))$$

$$\begin{aligned} P(Y \neq t) &= P(Y = +, t = -) + P(Y = -, t = +) \\ &= \left(P(+++) + 3P(++-)\right) P(t = -) + \left(P(---) + 3P(+--)\right) P(t = +) \\ &= \left[\left(\frac{2}{3}\right)^3 + 3\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)\right]\left(\frac{1}{3}\right) + \left[\left(\frac{1}{3}\right)^3 + 3\left(\frac{2}{3}\right)\left(\frac{1}{3}\right)^2\right]\left(\frac{2}{3}\right) \\ &= \frac{34}{81} \approx 42\% \end{aligned}$$

3.2 (b)

When the probability of "-" is 1/10, then the probability of "+" is 9/10. Therefore

$$\begin{aligned} P(Y \neq t) &= \left(P(+++) + 3P(++-)\right) P(t = -) + \left(P(---) + 3P(+--)\right) P(t = +) \\ &= \left[\left(\frac{9}{10}\right)^3 + 3\left(\frac{9}{10}\right)^2\left(\frac{1}{10}\right)\right]\left(\frac{1}{10}\right) + \left[\left(\frac{1}{10}\right)^3 + 3\left(\frac{9}{10}\right)\left(\frac{1}{10}\right)^2\right]\left(\frac{9}{10}\right) \\ &= 0.1224 = 12.24\% \end{aligned}$$

4 Decision Tree

4.1 (a)

Code the gender `"F"` as `1` and `"M"` as `0`. Similarly code the preference `"B"` as `1` and `"H"` as `0`.

[1]:
```
import pandas as pd
import numpy as np

data = np.array([[  1, 100,   1],
        [  0, 150,   1],
        [  1,  80,   1],
        [  0,  75,   1],
        [  1,  90,   0],
        [  1,  85,   0],
```

```
       [  0,  85,   0],
       [  1,  70,   0]])

data = pd.DataFrame(data)
data.columns = ["Gender", "Income", "Preference"]

data.sort_values("Income")
```

[1]:

	Gender	Income	Preference
7	1	70	0
3	0	75	1
2	1	80	1
5	1	85	0
6	0	85	0
4	1	90	0
0	1	100	1
1	0	150	1

So sammy will need to consider 4 values of a. They are $75, 85, 100, 150$. These are the values at which the preference changes.

4.2 (b)

$$\begin{aligned} Entropy(CarType) &= -p_+ \log_2 p_+ - p_- \log_2 p_- \\ &= -\left(\frac{4}{8}\right)\log_2\left(\frac{4}{8}\right) - \left(\frac{4}{8}\right)\log_2\left(\frac{4}{8}\right) \\ &= 1 \end{aligned}$$

4.3 (c)

$$Gain(S,a) = Entropy(S) - \sum_{v \in values(a)} \frac{|S_v|}{S} Entropy(S_v)$$

[2]:
```
def calculate_entropy(y):
    p_plus = sum(y)/len(y)
    p_minus = 1 - p_plus
    if ((p_plus == 1) or (p_minus == 1)):
        return 0
    else:
        entropy = -p_plus*np.log2(p_plus) - p_minus*np.log2(p_minus)
        return(entropy)

def calculate_gain(init_entropy, data, thresholds):
    gains = []
    gender_m = calculate_entropy(data[data["Gender"] == 0]["Preference"])
    gender_f = calculate_entropy(data[data["Gender"] == 1]["Preference"])
    prop_f = sum(data["Gender"])/len(data["Gender"])
```

```
    prop_m = 1-prop_f
    gender_gain = init_entropy-(prop_m*gender_m + prop_f*gender_f)
    gains.append(gender_gain)

    for a in thresholds:
        less_a = calculate_entropy(data[data["Income"] < a]["Preference"])
        greater_a = calculate_entropy(data[data["Income"] >= a]["Preference"])
        prop_less = sum(data["Income"] < a)/len(data["Income"])
        prop_greater = sum(data["Income"] >= a)/len(data["Income"])
        a_gain = init_entropy-(prop_less*less_a +prop_greater*greater_a)
        gains.append(a_gain)
    index = ["Gender"] + thresholds
    gains_df = pd.DataFrame(gains, index=index)
    gains_df.columns = ["Gain"]
    return(gains_df)
```

[3]:
```
root_gain = calculate_gain(1, data, [75, 85, 100, 150])
root_gain
```

[3]:

	Gain
Gender	0.048795
75	0.137925
85	0.048795
100	0.311278
150	0.137925

[4]:
```
root_gain.loc[root_gain.idxmax()]
```

[4]:

	Gain
100	0.311278

So splitting on annual income with $a = 100$ is the optimal root node. After splitting with this value, the group with $AI \geq 100$ only has 2 members, both which prefer `B`. Therefore we do not need to further split this group. For the $AI < 100$ group,

[5]:
```
data_100 = data[data["Income"] < 100]
init_100 = calculate_entropy(data_100["Preference"])
level2_gain = calculate_gain(init_100, data_100, [75, 85])
level2_gain
```

[5]:

	Gain
Gender	0.044110
75	0.109170
85	0.459148

[6]:
```
level2_gain.loc[level2_gain.idxmax()]
```

[6]:
```
        Gain
85  0.459148
```

On level 2, we split on $a = 85$. As with the previous split, the group with $a \geq 85$ has 3 members, all of which prefer `H`. Therefore, there is no further splitting for this group. For the other group,

[7]:
```python
data_85 = data_100[data_100["Income"] < 85]
init_85 = calculate_entropy(data_85["Preference"])
level3_gain = calculate_gain(init_85, data_85, [75])
level3_gain
```

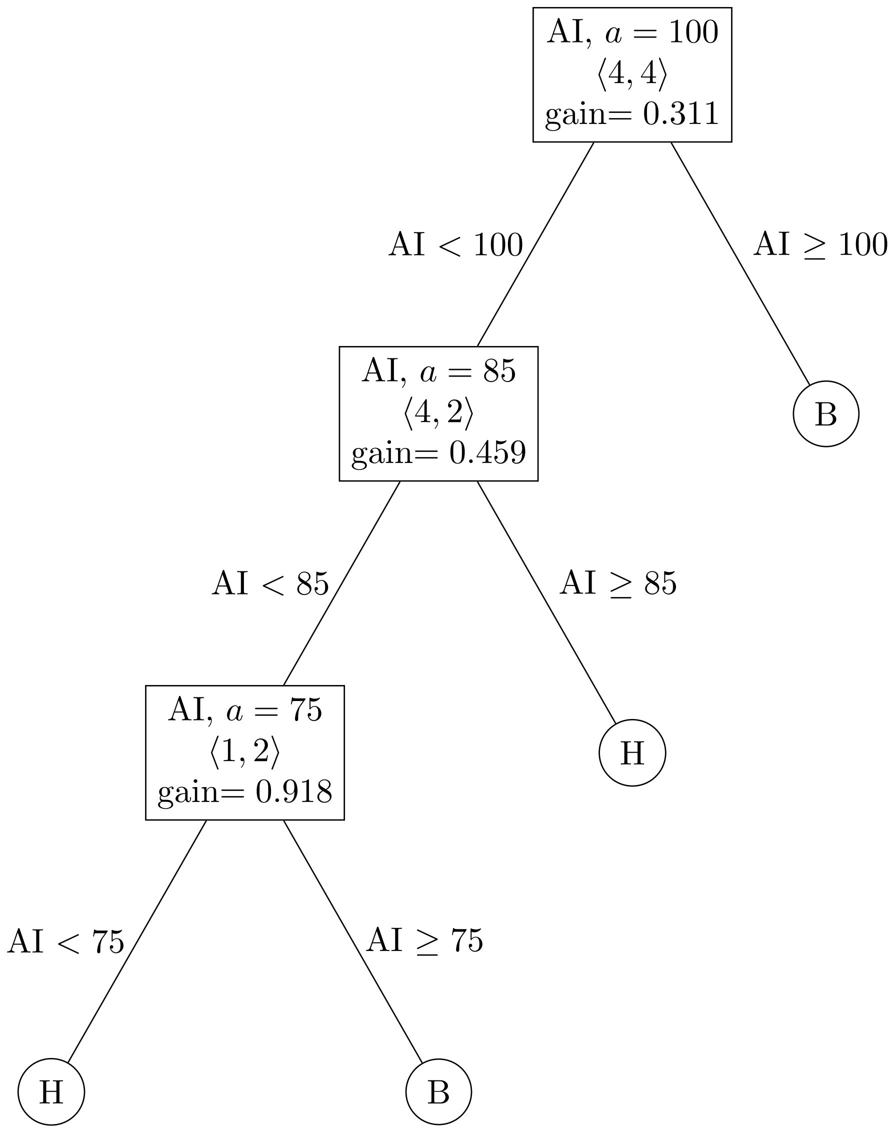
[7]:
```
            Gain
Gender  0.251629
75      0.918296
```

[8]:
```python
level3_gain.loc[level3_gain.idxmax()]
```

[8]:
```
        Gain
75  0.918296
```

So on level 3, we split on $a = 75$. This split separates the data into 2 groups, in which the the members of the groups prefer `H` or `B`. So this is the last level and we are done.

(d)


AI, $a = 100$
$\langle 4, 4 \rangle$
gain= 0.311
AI < 100
AI ≥ 100
AI, $a = 85$
$\langle 4, 2 \rangle$
gain= 0.459
B
AI < 85
AI ≥ 85
AI, $a = 75$
$\langle 1, 2 \rangle$
gain= 0.918
H
AI < 75
AI ≥ 75
H
B