Name: Harris Spahic

Pledge: "I pledge my honor I have abided by the Stevens Honor System."

- Harris Spahic

#1)* $P(CS) = P(S_1) \times P(CS \mid S_1)$
$+ P(S_2) \times P(CS \mid S_2)$
$+ P(S_3) \times P(CS \mid S_3)$

$$= 0.2 \times 6/20 + 0.2 \times 10/20 + 0.6 \times 6/20$$

$$= 0.8(6/20) + 0.2(10/20)$$

$$= \frac{4.8}{20} + \frac{2}{20} = \frac{6.8}{20} = \frac{34}{100} = 0.34$$

1.2) $$P(S3 \mid Stat) = \frac{P(S3) \times P(Stat \mid S3)}{P(Stat)}$$

$$= \frac{0.6(6/20)}{24/60} = \frac{3.6 \times 60}{24 \times 20} = \frac{10.8}{24} = 0.45$$

#2) Since the weights are normally distributed, the Likelihood function is

Probability Mass Function

$$= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x_i-\mu)^2}{2\sigma^2}\right)$$

$$\rightarrow LE(\mu,\sigma^2) = \prod_{i=1}^{n} f(x_i;\mu,\sigma^2)$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(\sum_{i=1}^{n} \frac{-(x_i-\mu)^2}{2\sigma^2}\right)$$

$$= \left(\frac{1}{\sigma}\right)^n \cdot \left(\frac{1}{2\pi}\right)^{n/2} \exp\left(\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(-(x_i-\mu)^2\right)\right)$$

(why can't this be "minimum LE"

2.2) To find the MLE(μ) & MLE(σ^2 we must take $d/d\theta$, where θ is the parameter in question, and set it equal to 0.

Looking at our LE (μ, σ) from #2, $d/d\theta$ is hard to compute. But notice $\ln(x)$ is a continously increasing function, hence the max of $\ln(MLE(\theta))$ and max $MLE(\theta)$ both at same θ.

→ $MLE(\mu) \Rightarrow \frac{d}{d\mu} \ln(LE(\mu, \sigma^2)) = 0$

$$= \frac{d}{d\mu}\left(-n\ln(\sigma) - \frac{n}{2}\ln(2\pi) + \left(\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}\left(-(x_i-\mu)^2\right)\right)\right)\right)$$

$$= 0 + 0 + \frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} -2(x_i-\mu)\cdot -1\right)$$

$$= \frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} x_i - n\mu\right) = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i - n\mu = 0$$

$$\Rightarrow \frac{\sum_{i=1}^{n} x_i}{n} = \mu_{ML}$$

→ $MLE(\sigma^2) \Rightarrow (MLE(\theta_2))$ where $\theta_2 = \sigma^2$

$$\Rightarrow LE(\mu, \theta_2) = \left(\frac{1}{\theta}\right)^{\frac{n}{2}} \cdot \left(\frac{1}{2\pi}\right)^{n/2} \exp\left(\frac{1}{2\theta_2}\sum_{i=1}^{n}\left(-(x_i-\mu)^2\right)\right)$$

$$\Rightarrow MLE(\theta_2) = \frac{d}{d\theta_2}\ln\left(LE(\mu, \theta_2)\right)$$

$$= \frac{d}{d\theta_2}\left(-\frac{n}{2}\ln(\theta) - \frac{n}{2}(2\pi) + \left(\frac{1}{2\theta_2}\sum_{i=1} -(x_i-\mu)\right)\right)$$

$$= -\frac{n}{2\theta_2} - 0 + \frac{1}{2\theta_2^2}\cdot\sum_{i=1}^{n}(x_i-\mu)^2 = 0$$

$$\Rightarrow -n\theta_2 + \sum_{i=1}^{n}(x_i-\mu)^2 = 0$$

Since $\mu = \bar{x}$ by MLE(μ)

$$\Rightarrow \sum_{i=1}^{n}(x_i-\mu)^2 = n\theta_2$$

$$\Rightarrow \frac{\sum_{i=1}^{n}(x_i-\mu_{ML})^2}{n} = \theta_2 = \sigma^2_{ML}$$

Thus according to MLE(μ).

$$\hat{\mu} = \frac{112+120+131+126+145+158+157+136+148+176}{10}$$

$$= \frac{1409}{10} = 140.9$$

$$\tilde{\sigma}^2 = \frac{(112-140.9)^2 + \cdots + (176-140.9)^2}{10}$$

$$= 346.69$$

3.1) $L.E(q) = \prod_{i=1}^{n} f(x_i; q)$

where $f(x)$ = Probability Mass Function

Thus,

$$L.E(\mu, \sigma^2) = f(x_1; q)^{n_1} \times f(x_2; q)^{n_2} \times f(x_3; q)^{n_3} \times f(x_4; q)^{n_4}$$

where n_x are the # of occurances of x_i in our sample.

$$= \left(\frac{2q}{3}\right)^{n_1} \left(\frac{q}{3}\right)^{n_2} \left(\frac{2(1-q)}{3}\right)^{n_3} \left(\frac{1-q}{3}\right)^{n_4}$$

Our sample

(1, 1, 2, 2, 2, 3, 3, 3, 4, 4)

$$\rightarrow \left(\frac{2q}{3}\right)^2 \left(\frac{q}{3}\right)^3 \left(\frac{2(1-q)}{3}\right)^3 \left(\frac{1-q}{3}\right)^2$$

$$= \frac{2^5 q^5 (1-q)^5}{3^{10}}$$

3.2) $LE(q) = \left(\frac{2q}{3}\right)^{n_1} \left(\frac{q}{3}\right)^{n_2} \left(\frac{2(1-q)}{3}\right)^{n_3} \left(\frac{1-q}{3}\right)^{n_4}$

→ To find MLE(q),

$d/dq\, LE(q) = 0$ & since $\ln(x)$ is an continuously increasing function, the value of q that maximizes $LE(q)$ is the same as $\ln(LE(q))$.

$$\rightarrow d/dq \ln\left(\left(\frac{2q}{3}\right)^{n_1} \left(\frac{q}{3}\right)^{n_2} \left(\frac{2(1-q)}{3}\right)^{n_3} \left(\frac{1-q}{3}\right)^{n_4}\right) = 0$$

$$\rightarrow d/dq \left(n_1 \ln\left(\frac{2q}{3}\right) + n_2 \ln\left(\frac{q}{3}\right) + n_3 \ln\left(\frac{2(1-q)}{3}\right) + n_4 \ln\left(\frac{1-q}{3}\right)\right) = 0$$

All constants will come out of $\ln(\sim)$ & be eliminated by d/dq

$$\rightarrow \frac{d}{dq}\left(n_1 \ln(q) + n_2 \ln(q) + n_3 \ln(1-q) + n_4 \ln(1-q)\right)$$

$$= \frac{n_1}{q} + \frac{n_2}{q} - \frac{n_3}{(1-q)} - \frac{n_4}{(1-q)} = \frac{n_1+n_2}{q} - \frac{n_3+n_4}{(1-q)}$$

$$= \frac{n_1+n_2 - q(n_1+n_2+n_3+n_4)}{q(1-q)} = 0$$

$$\rightarrow \frac{n_1+n_2}{(n_1+n_2+n_3+n_4)} = q_{ML}$$

Since $n_1 = 2$
$n_2 = 3$
$n_3 = 3$
$n_4 = 2$

$$q_{ML} = 1/2$$

Notice $q = 0$ & $q = 1$, then
$LE(q) = 0$.

Thus $q = \frac{1}{2}$ is the most likely value of q given the sample.

---

4) Recall:

$$P(\theta | x) = \frac{P(x|\theta) P(\theta)}{P(x)}$$

which implies

$$P(\theta | X) \propto P(X|\theta) P(\theta)$$

Notice: $P(x|\theta)$ is equivalent to the likelihood function.

Thus finding the maximum posterior function ($P(\theta | X)$), is equivalent to maximizing the likelihood $\times P(\theta)$.

Then given
N input values, $x = (x_1, x_2, \dots, x_n)^T$
and target values $y = (y_1, \dots, y_N)^T$;

we estimate the target using $f(x,w)$. Assuming target variables drawn from a Guassian Distribution

$$P(y \mid x, w, \beta) = N(y \mid f(x,w), \beta^{-1})$$

& a prior Guassian Distribution for w

$$P(w \mid \alpha) = \left(\frac{\alpha}{2\pi}\right)^{(M+1)/2} \exp\left(-\frac{\alpha}{2} w^T w\right)$$

We procede to find the MAP using
$\max(P(y \mid x, w, \beta) \times p(w \mid \alpha))$

$$p(w \mid x, y, \alpha, \beta) \propto P(y \mid x, w, \beta)\, p(w \mid \alpha)$$

$$p(w \mid x, y, \alpha, \beta) \propto \prod_{n=1}^{N} N(y \mid f(x_n, w), \beta^{-1})\, p(w \mid \alpha)$$

$$\rightarrow \left((\beta)^{N/2} \left(1/2\pi\right)^{N/2} \exp\left(\frac{1}{2\beta^{-1}} \sum_{n=1}^{N} \left(-(f(x_n,w) - y_n)^2\right)\right)\right) \left(\frac{\alpha}{2\pi}\right)^{\frac{(M+1)}{2}} \exp\left(-\frac{\alpha}{2} w^T w\right)$$

maximize log in terms w

$$d/dw \left(N/2 \ln \beta - N/2 \ln 2\pi - \frac{\beta}{2} \sum_{n=1}^{N} (f(x_n,w) - y_n)^2 + \frac{M+1}{2} \ln\left(\frac{\alpha}{2\pi}\right) - \frac{\alpha}{2} w w^T \right)$$

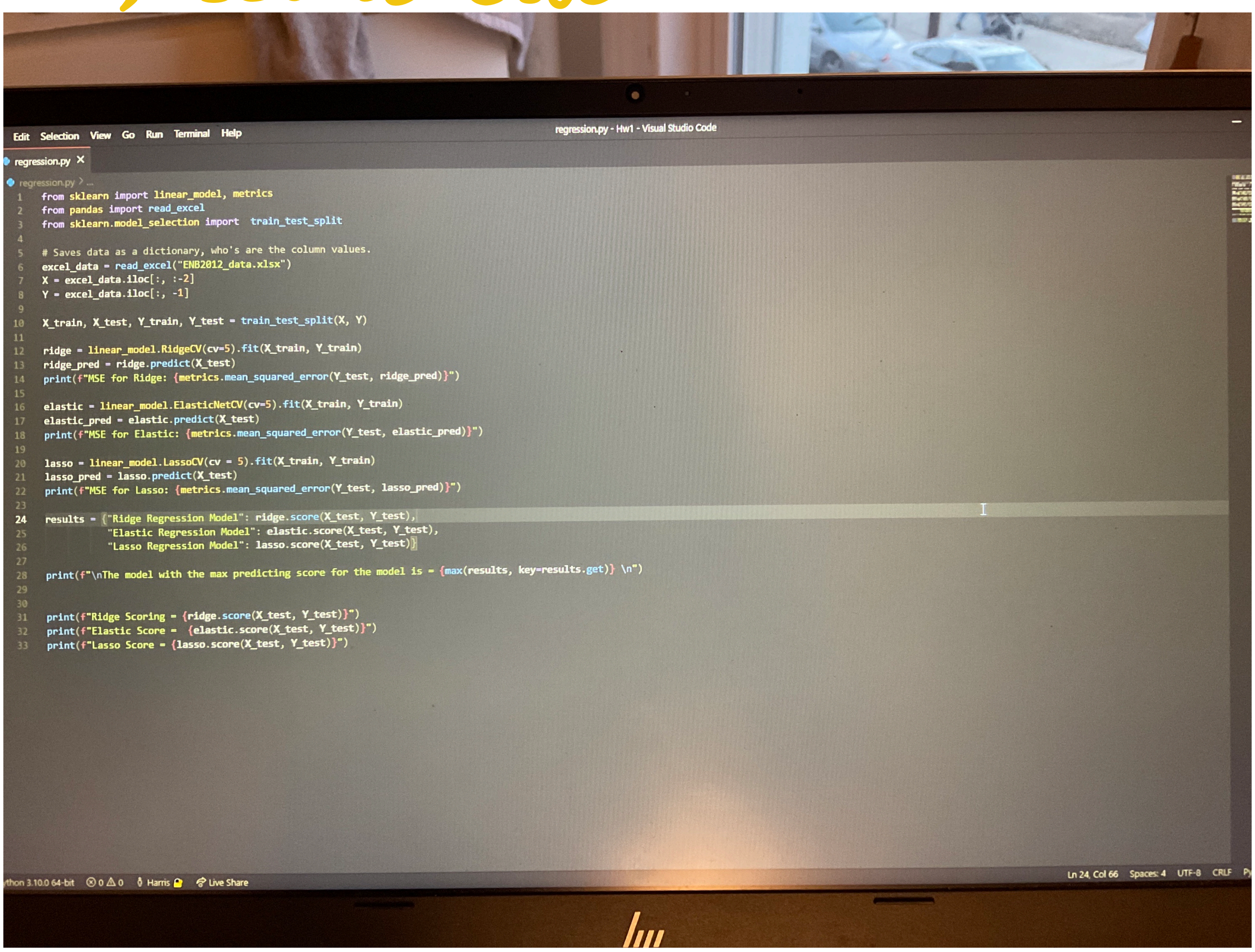
$$= -\frac{\beta}{2}\sum_{n=1}^{N}\left(f(x_n,w)-y_n\right)^2-\frac{\alpha}{2}ww^T$$

Notice maximizing log in w is equivalent to minimizing the negative log in w.

$$\rightarrow = \frac{d}{dw}\left(\frac{\beta}{2}\sum_{n=1}^{N}\left(f(x_n,w)-y_n\right)^2\right)+\frac{\alpha}{2}ww^T\Big) = 0$$

Or the minimizing of the negative sum-of-squares error function.

#5) Source Code



```
from sklearn import linear_model, metrics
from pandas import read_excel
from sklearn.model_selection import  train_test_split

# Saves data as a dictionary, who's are the column values.
excel_data = read_excel("ENB2012_data.xlsx")
X = excel_data.iloc[:, :-2]
Y = excel_data.iloc[:, -1]

X_train, X_test, Y_train, Y_test = train_test_split(X, Y)

ridge = linear_model.RidgeCV(cv=5).fit(X_train, Y_train)
ridge_pred = ridge.predict(X_test)
print(f"MSE for Ridge: {metrics.mean_squared_error(Y_test, ridge_pred)}")

elastic = linear_model.ElasticNetCV(cv=5).fit(X_train, Y_train)
elastic_pred = elastic.predict(X_test)
print(f"MSE for Elastic: {metrics.mean_squared_error(Y_test, elastic_pred)}")

lasso = linear_model.LassoCV(cv = 5).fit(X_train, Y_train)
lasso_pred = lasso.predict(X_test)
print(f"MSE for Lasso: {metrics.mean_squared_error(Y_test, lasso_pred)}")

results = {"Ridge Regression Model": ridge.score(X_test, Y_test),
           "Elastic Regression Model": elastic.score(X_test, Y_test),
           "Lasso Regression Model": lasso.score(X_test, Y_test)}

print(f"\nThe model with the max predicting score for the model is = {max(results, key=results.get)} \n")


print(f"Ridge Scoring = {ridge.score(X_test, Y_test)}")
print(f"Elastic Score =  {elastic.score(X_test, Y_test)}")
print(f"Lasso Score = {lasso.score(X_test, Y_test)}")
```

Evaluation Result