MAD Assignment 2

Ask Jensen

6. december 2021

Indhold

1 Problem 1

1.1 (a)

$$\mathscr{L} = \frac{1}{N}\sum_{n=1}^{N} \alpha_n(\mathbf{w}^T\mathbf{x}_n - t_n)^2$$

$$= \frac{1}{N}(\mathbf{Xw} - t)^T A(\mathbf{Xw} - t)$$

$$= \frac{1}{N}(\mathbf{Xw})^T - t^T(A\mathbf{Xw} - At)$$

$$\frac{1}{N}(\mathbf{Xw})^T A\mathbf{Xw} - \frac{1}{N}(\mathbf{Xw})^T At - \frac{1}{N}A\mathbf{Xw}t^T + \frac{1}{N}Att^T$$

$$= \frac{1}{N}\mathbf{w}^T\mathbf{X}^T A\mathbf{Xw} - \frac{2}{N}\mathbf{w}^T\mathbf{X}^T At + \frac{1}{N}t^T At$$

using case 4 and case 1 from the table 1.14 when differentiating I get:

$$\frac{\partial\mathscr{L}}{\partial w} = 2\mathbf{X}^T\mathbf{AXw} - 2\mathbf{X}^T\mathbf{A}t = 0$$

+2 and -2 cancel out leaving me with

$$\mathbf{X}^T\mathbf{AXw} - \mathbf{X}^T\mathbf{A}t = 0$$

$$\mathbf{X}^T\mathbf{AXw} = \mathbf{X}^T\mathbf{A}t$$

multiplying both sides with the identity matrix

$$\mathbf{Iw} = (\mathbf{X}^T\mathbf{AX})^{-1}\mathbf{X}^T\mathbf{A}t$$

Multiplying the vector **w** with the Identity matrix, will simply return the vector **w** thus the result is

$$\hat{\mathbf{w}} = (\mathbf{X}^T\mathbf{AX})^{-1}\mathbf{X}^T\mathbf{A}t$$

1.2 (b)

1. What do you expect to happen? I expect that the regression will have an overall better fit.
2. What do you observe? That the predictions has been scattered more than previously. However, smaller values now fit the regression line better. But points that fitted better before, might become outliers.
3. Do the additional weights have an influence on the outcome The weights has as mentioned before, a positive influence on points with a low value, however a negative influence on values that might fit the regression well from the beginning.

Source code for ex1b

```
# fit linear regression model using all features
print("Linear fit:")
model_all = linweighreg.LinearRegression()
model_all.fit(X_train, t_train)
all_values = model_all.w

# Prints all the weights from fit()
for i in range(len(all_values)):
    print("\tw%i : %s" %(i, model_all.w[i]))

# compute corrospondingp predictions in boston_test set
pred_all = model_all.predict(X_test)

# Plot the regression
x_single = t_test
y_single = pred_all
plt.title('True House Prices vs. Weighted Estimates')
plt.xlabel('True House Prices')
plt.ylabel('Estimates')
plt.plot(t_test, t_test, c='r')
plt.scatter(x_single,y_single)
plt.show()
```

linweighreg.py

```
def fit(self, X, t):
"""
Fits the linear regression model.

Parameters
----------
X : Array of shape [n_samples, n_features]
t : Array of shape [n_samples, 1]
```

```
"""
# TODO: YOUR CODE HERE
n = X.shape[0]
X = np.array(X).reshape((n, -1))
t = np.array(t).reshape((n, 1))

# prepend a column of ones
ones = np.ones((X.shape[0], 1))
X = np.concatenate((ones, X), axis=1)

A = np.identity(len(t))
A_pow = A*(t**2)

# compute weights (solve system)
a = np.dot(X.T, A_pow)
b = np.dot(a, X)
c = np.dot(a,t)

self.w = np.linalg.solve(b,c)
```

2 Problem 2

Some of the code has been implemented but is not finished. See the following source code

Source code for running LOOCV

```
# Reads the data from the file
raw = np.genfromtxt('men-olympics-100.txt', delimiter=' ')
transposed = raw.T

# Extract the first "row", (index 1, since the array is 0-indexed) from raw
OL_year = raw[:,0].T

# Extract the second "row", (index 1, since the array is 0-indexed) from raw
OL_run_times = raw[:,1].T

# Create lamda values for LOOCV
lambda_values = np.logspace(-8, 0, 100, base=10)
print("lambda shape:", lambda_values.shape)
#print(lambda_values)

#Olympic years
x = OL_year

#First place values
```

```
y = OL_run_times

model_all = linweighreg.LinearRegression()
model_all.fit(x,y)
all_weights = model_all.w
print("---------------------------------------")
for i in range(len(all_weights)):
    print("\tw%i : %s" %(i, model_all.w[i]))
print("---------------------------------------")


N = len(y)

def RMSE(t,tp):
    res = np.sqrt(np.square(np.subtract(t,tp)).mean())
    print(res)
    return(res)

for i in lambda_values:
    model_LOOCV = linweighreg.LinearRegression()
    LOOCV_res = model_LOOCV.fit_LOOCV(x, y, i, N)
    loss = RMSE(LOOCV_res.w, y)
    print("lam=%.10f and loss=%.10f" % (lambda_values, loss))



plt.title('Mens 100m sprint results')
plt.xlabel('Olympic year')
plt.ylabel('1st place 100m track time')
plt.scatter(x, y)

xplot=np.linspace(1896,2008,100)
poly =np.polyfit(x,y,1)

xplot=np.linspace(1896,2008,100)
poly =np.polyfit(x,y,1)

print("values from polynomial fit:", poly)

yplot = poly[1]+poly[0]*(xplot)
plt.plot(xplot,yplot, c='r')
```

Code for the fit_LOOCV.

```
def fit_LOOCV(self, X, t, lamda, N):
# Create identity matrix
n = X.shape[0]
```

```
X = np.array(X).reshape((n, -1))
t = np.array(t).reshape((n, 1))
idm = np.identity(X.shape[1])

# X^T * X
xTx = np.dot(X.T, X)
print("xTx shape:", xTx.shape)

# X^T * t
xTt = np.dot(X.T, t)
print("xTt shape:", xTt.shape)

# N * Lamda * identity_matrix
INLamda = N * lamda * idm
print("Lambda shape",INLamda.shape)

fst_block = xTx + INLamda
print("fst shape:", fst_block.shape)

#coef = np.dot(np.dot(np.linalg.inv(fst_block), X.T), t)
self.w = np.linalg.solve(fst_block,xTt)
```

.

3 Problem 3

3.1 (a)

The a PDF is the derivative of a CDF.
The function $e^{-\beta x^{\alpha}}$ is a composition of f(g(x)). Thus the after applying the chainrule and deriving where $g(x) = -\beta x^{\alpha}$:

$$\frac{d}{dx} f(g(x)) = \frac{d}{dx}(-\beta x^{\alpha} \cdot e^{g(x)}) = -\beta \alpha x^{\alpha-1} \cdot e^{-\beta x^{\alpha}}$$

Thus the PDF is:

$$f(x) \begin{cases} -\beta \alpha x^{\alpha-1} \cdot e^{-\beta x^{\alpha}} & \text{if} \quad x > 0 \\ 0 & \text{otherwise} \end{cases}$$

3.2 (b)

In this question there are two subquestions.

1. What is the probability that the chip works longer than four years?
2. What is the probability that the chip stops working in the time interval [5; 10] years?

I'm given two values for α and β stubstituting these values into the original function I have:

$$f(x) = 1 - e^{-\frac{1}{4}x^2} = 1 - e^{-\frac{x^2}{4}}$$

To answer the first question I can simply calculate f(5). Since I'm asked to answer the probability for the chip to live *more* than four years. It is important to note that when the value = 1 the chip is dead. Otherwise, the chip would become better over time. $f(5) = 0.998069$ This means, that after 5 years, the chip is almost certanly dead.

To calculate the probability that the chip will die somewhere between the interval $[5;10]$ I can do it like so:

$$F(10) - F(5) = 0.0183156$$

This means that the chip has $\approx 1.83\%$ chance of surviving. So to make this more readble I can say:

$$1 - (F(10) - F(5)) \approx 98.16\% \quad \text{chance of death}$$

in the interval from 5 to 10 years.

3.3 (c)

Finding the median of the function. I can simply set it equal to $\frac{1}{2}$ and solve it for x.

$$1 - e^{-\beta x^\alpha} = \frac{1}{2}$$

simple rewriting

$$1 - \frac{1}{2} = e^{-\beta x^\alpha}$$

multiplying with $ln()$ to remove the exponent

$$ln(\frac{1}{2}) = -\beta x^\alpha$$

dividing with $-\beta$

$$\frac{ln(\frac{1}{2})}{-\beta} = x^\alpha$$

$ln(\frac{1}{2})$ is the same as $-ln(2)$ so rewriting and removing the minuses

$$\frac{ln(2)}{\beta} = x^\alpha$$

taking $ln(x^\alpha)$ to move the α

$$\ln(\frac{ln(2)}{\beta}) = \alpha \cdot ln(x)$$

moving alpha

$$\frac{\ln(\frac{ln(2)}{\beta})}{\alpha} = ln(x)$$

We already know that in order to remove the exponenet, we can use ln. so to remove ln I can add back the exponent, which gives me the answer

Answer:

$$e^{\frac{\ln(\frac{ln(2)}{\beta})}{\alpha}} = x$$

4 Problem 4

4.1 (a)

Since Peter dosn't have any prior conviction and the exercise is to calculate X_{speak} I will only be looking at point 2),4) and 6) given with the exercise.
I will be substituting the with the values given in the above points from right to left, since this gives to most readability in context to the exercise.

$P(\texttt{Conviction}) = P(\texttt{Conviction}|\texttt{Court})P(\texttt{Court} \cap \texttt{Speaks} \cap \texttt{NC})$
$P(\texttt{Conviction}) = P(\texttt{Conviction}|\texttt{Court}) \cdot 0.002$
$P(\texttt{Conviction}) = 0.5 \cdot 0.002 = 0.001$
Peters conviction $= 5\,\texttt{years} \cdot 0.75 \; = \; 3.75\,\texttt{years}$

$$P(X = x) \begin{cases} 0.001 & \text{for} \quad x = 3.75 \text{ years} \\ 0.998 & \text{for} \quad x = 0 \quad \text{(no prison time)} \end{cases}$$

$EX_{speaks} \; = \; 0.001 \cdot 3.75 + 0.998 \cdot 0 = 0.00375$

4.2 (b)

I will be using the same procedure as in 4a. With the exception that now Peter will remain silent meaning I will be using point 1), 4), 5) and 6)

$P(\texttt{Conviction}) = P(\texttt{Conviction}|\texttt{Court})\; P(\texttt{Court} \cap \texttt{Silent} \cap \texttt{NC})$
$P(\texttt{Conviction}) = P(\texttt{Conviction}|\texttt{Court}) \cdot 0.002$
$P(\texttt{Conviction}) = (1 - (0.5 \cdot \frac{1}{4})) \cdot 0.001$
$P(\texttt{Conviction}) = 0.875 \cdot 0.001 = 0.000875$
Peters conviction $= 5\,\texttt{years}$

$$P(X = x) \begin{cases} 0.000875 & \text{for} \quad x = 5 \text{ years} \\ 0.999125 & \text{for} \quad x = 0 \quad \text{(no prison time)} \end{cases}$$

$EX_{silent} \; = \; 0.000875 \cdot 5 + 0.999125 \cdot 0 = 0.004375$

4.3 (c)

Brian talks and is previously convicted. I will be using point 1),3),4),5) and 6) in order to compute Y_{speaks} and Y_{silent} for Brian

$P(\texttt{Conviction}) = P(\texttt{Conviction}|\texttt{Court})\ P(\texttt{Court} \cap \texttt{Silent} \cap \texttt{C})$
$P(\texttt{Conviction}) = P(\texttt{Conviction}|\texttt{Court}) \cdot 0.005$
$P(\texttt{Conviction}) = (1-0.1) \cdot 0.005 = 0.0045$
Brians conviction $= 5\ \texttt{years} \cdot 0.75\ =\ 3.75\ \texttt{years}$

$$P(Y=y) \begin{cases} 0.0045 & \text{for} \quad y = 3.75 \text{ years} \\ 0.9955 & \text{for} \quad y = 0 \quad \text{(no prison time)} \end{cases}$$

$EY_{speaks}\ =\ 0.0045 \cdot 3.75 + 0.9955 \cdot 0 = 0.016875$

$P(\texttt{Conviction}) = P(\texttt{Conviction}|\texttt{Court})\ P(\texttt{Court} \cap \texttt{Silent} \cap \texttt{C})$
$P(\texttt{Conviction}) = P(\texttt{Conviction}|\texttt{Court}) \cdot 0.001$
$P(\texttt{Conviction}) = (1-(0.5 \cdot \frac{0.1}{4})) \cdot 0.001 = 0.0045$
$P(\texttt{Conviction}) = 0.975 \cdot 0.001 = 0.000975$
Brians conviction $= 5\ \texttt{years} \cdot 0.75\ =\ 3.75\ \texttt{years}$

$$P(Y=y) \begin{cases} 0.000975 & \text{for} \quad y = 3.75 \text{ years} \\ 0.999025 & \text{for} \quad y = 0 \quad \text{(no prison time)} \end{cases}$$

$EY_{silent}\ =\ 0.000975 \cdot 3.75 + 0.999025 \cdot 0 = 0.016875$

Thus for Peter

$$\text{EX}_{\texttt{speak}} \quad < \quad \text{EX}_{\texttt{silent}}$$

and for Brian

$$\text{EY}_{\texttt{speak}} \quad > \quad \text{EY}_{\texttt{silent}}$$