Quiz 3 Solutions

CSE 4/574

Fall, 2024

Question 1

This problem deals with linear regression. Consider the following data set, the input variable (first column) indicates the amount spent by an individual on internet service and the output variable (second column) indicates the degree of satisfaction with his/her internet service.

Time Spent	Satisfaction Degree
11	6
18	8
17	10
15	4
9	9
5	6
12	3
19	5
22	2
25	10

We are interested in choosing the best weight vector among four candidates. The fourth candidate (w4) uses higher-order features - 1,x,x2 . For all others we assume that first component of the weight vector corresponds to the bias/intercept term. We assume that the variance for each y is same (0.01) for all choices of the weights. We assume a Gaussian prior on the weight vectors with 0 mean and a diagonal covariance matrix with 0.01 at every diagonal entry.

The candidates are:

$$
\begin{aligned}
w_1 &= (0.30, 0.50) \\
w_2 &= (5.75, 0.04) \\
w_3 &= (3.20, 0.20) \\
w_4 &= (8.75, -0.50, 0.02)
\end{aligned}
\tag{1}
$$

Given the following statements:

1. Among the first three weights, w2 will be the best estimate in terms of likelihood of the given data.
2. Among the first three weights, w3 will be the best estimate in terms of likelihood of the given data.
3. Among all four weights, w2 will be the best estimate in terms of likelihood of the given data.
4. Among all four weights, w4 will be the best estimate in terms of likelihood of the given data.
5. w3 has the highest posterior.
6. w4 has the highest posterior.
7. Ridge regression will always choose a non-linear mapping on features over the raw linear feature.

Which of the following are true?

Correct Choice

Statements 1 and 4 are true and Statement 7 is false.

Incorrect Choice 1

Statements 4 and 6 are true and Statement 7 is false.

Incorrect Choice 2

Statements 5 and 7 are false.

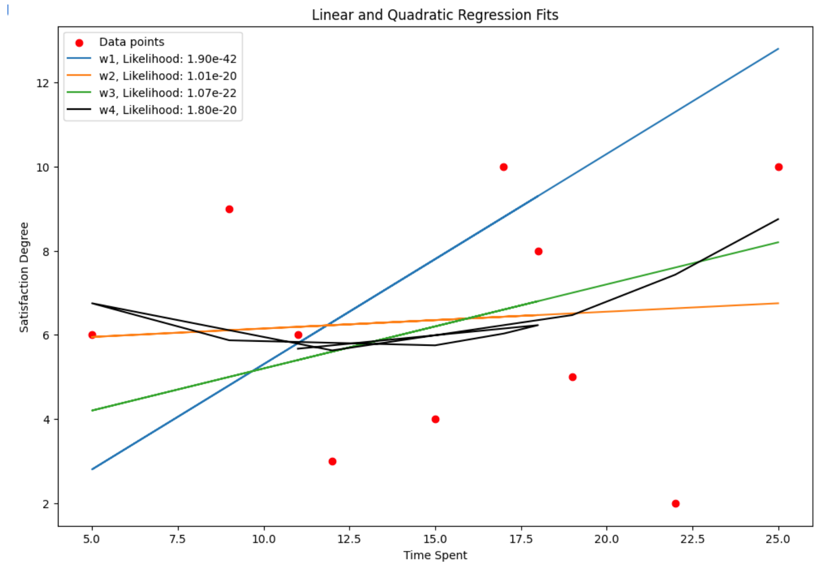
Incorrect Choice 3

Statements 2 and 3 are true.

Problem Explanation:

```
import numpy as np
import matplotlib.pyplot as plt

# Given data
time_spent = np.array([11, 18, 17, 15, 9, 5, 12, 19, 22, 25])
satisfaction_degree = np.array([6, 8, 10, 4, 9, 6, 3, 5, 2, 10])

# Given weight vectors
w1 = np.array([0.30, 0.50])
w2 = np.array([5.75, 0.04])
w3 = np.array([3.20, 0.20])
w4 = np.array([8.75, -0.50, 0.02])

# Error variance
sigma2 = 1 #using larger variance instead of 0.01 to show likelihood

# Models
def linear_model(x, w):
    return w[0] + w[1] * x

def quadratic_model(x, w):
    return w[0] + w[1] * x + w[2] * x**2

# Calculate likelihood of each model
def likelihood(y, y_pred, sigma2):
    return np.prod(np.exp(-0.5 * ((y - y_pred)**2 / sigma2)) / np.sqrt(2 * np.pi * sigma2))
```

```
# Predictions
y_pred_w1 = linear_model(time_spent, w1)
y_pred_w2 = linear_model(time_spent, w2)
y_pred_w3 = linear_model(time_spent, w3)
y_pred_w4 = quadratic_model(time_spent, w4)

# Likelihoods
likelihood_w1 = likelihood(satisfaction_degree, y_pred_w1, sigma2)
likelihood_w2 = likelihood(satisfaction_degree, y_pred_w2, sigma2)
likelihood_w3 = likelihood(satisfaction_degree, y_pred_w3, sigma2)
likelihood_w4 = likelihood(satisfaction_degree, y_pred_w4, sigma2)

# Plotting
plt.figure(figsize=(12, 8))
plt.scatter(time_spent, satisfaction_degree, color='red', label='Data points')
plt.plot(time_spent, y_pred_w1, label=f'w1, Likelihood: {likelihood_w1:.2e}')
plt.plot(time_spent, y_pred_w2, label=f'w2, Likelihood: {likelihood_w2:.2e}')
plt.plot(time_spent, y_pred_w3, label=f'w3, Likelihood: {likelihood_w3:.2e}')
plt.plot(time_spent, y_pred_w4, label=f'w4, Likelihood: {likelihood_w4:.2e}', color='black')
plt.title('Linear and Quadratic Regression Fits')
plt.xlabel('Time Spent')
plt.ylabel('Satisfaction Degree')
plt.legend()
plt.show()

print('Likelihoods:')
print('w1:', likelihood_w1)
print('w2:', likelihood_w2)
print('w3:', likelihood_w3)
print('w4:', likelihood_w4)
```




```
Likelihoods:
w1: 1.9044410885682692e-42
w2: 1.0066384654113185e-20
w3: 1.0698160457373624e-22
w4: 1.7993338765638248e-20
```

From the likelihood calculation, we know that statement 1 is true, while statement 2 is false. As w_4 has the highest likelihood among all weights, statement 3 is false while statement 4 is true.
Based on the equation in Fig. 1, we know that posterior is proportional to prior. As w_4 has weights that deviate much more than w_2, with 0.01 as the variance, w_4's has a much smaller prior probability. As w_2 and w_4 has likelihood in a similar magnitude, w_4's posterior is at least worse than that of w_2's. Hence, statement 6 is false. This leave the correct choice the only remaining choice.

- What is posterior of $\mathbf{w}$

$$p(\mathbf{w}|\mathcal{D}) \propto \prod_i \mathcal{N}(y_i|\mathbf{w}^\top \mathbf{x}_i, \sigma^2)p(\mathbf{w})$$

Figure 1: Equation for posterior calculation.

Question 2

Consider the problem of learning a linear regression model with the following form of a regularized error function.

$$\widehat{\mathbf{w}} = \arg\min_{\mathbf{w}} \frac{1}{2}\sum_{i=1}^{N}(y_i - \mathbf{w}^\top \mathbf{x}_i)^2 + \frac{1}{2}\alpha\|\mathbf{w}\|^2$$

The above equation corresponds to ridge regression. In this problem, you need to find the best weight vector (w) for the training data below according to the above regularized loss function. Make sure that you include the bias term in the input data. The last element in w corresponds to the bias term.

i	x_i	y_i
1	-1.0	1.0
2	0.0	2.0
3	1.0	3.0

Assume that the regularization parameter $= 1$. Choose the optimal weight below:

Correct Choice

$w = [0.67, 1.5]$

Incorrect Choice 1

$w = [1.5, 0.67]$

Incorrect Choice 2

$w = [1.5, 0.5]$

Incorrect Choice 3

$w = [0.5, 2.0]$

Question Explanation:
By drawing the 3 data points in the 2D coordinate, we see that the slope = 1, intercept = 2 without any weight penalty. By applying ridge regression, we effectively lower the sensitivity of the change of the predicted values due to change in input value.
By calculating the loss according to the equation, we get:

$$\begin{aligned} w &= [0.67, 1.5]\,, loss \approx 2.5 \\ w &= [1.5, 0.67]\,, loss \approx 4.9 \\ w &= [1.5, 0.5]\,, loss \approx 5.5 \\ w &= [0.5, 2.0]\,, loss \approx 3.4 \end{aligned} \tag{2}$$

Hence, the correct answer is $w = [0.67, 1.5]$