Concept Quiz Over Week 2 Material

Due Oct 15 at 11:59pm **Points** 1 **Questions** 10
Available Oct 11 at 12am - Oct 15 at 11:59pm **Time Limit** None

Score for this survey: **1** out of 1
Submitted Oct 15 at 2:50pm
This attempt took 204 minutes.

Question 1

Assuming we are using a Bayesian approach to fitting the parameters of a model to some data, match the names of the following distributions to their written versions.

ou Answered	**Posterior**	P(parameters | data)
ou Answered	**Prior**	P(parameters)
ou Answered	**Likelihood**	P(data | parameters)

Posterior = P(parameters | data)

Prior = P(parameters)

Likelihood = P(data|parameters)

Question 2

What role does a prior play in MAP inference?

Your Answer:

The prior acts as a form of regularization, It can help prevent overfitting by biasing the model towards parameter values that are more probable based on prior knowledge, which can be especially useful when you have limited data. If you have a strong prior belief, it can have a substantial influence on the posterior, particularly when the likelihood (the fit of the data) is weak. Thus the prior allows you to combine your prior knowledge or beliefs with new data to make more informed and regularized inferences about the model parameter.

A prior is a way for us to encode beliefs about the values of our parameters _before_ seeing any data.

Question 3

If a prior distribution A is a conjugate prior to likelihood distribution B, what can I say about the posterior distribution B*A?

Your Answer:

Then posterior distribution B*A will belong to the same family of probability distributions as the prior distribution A. In other words, the posterior distribution will have the same form or type as the prior distribution. This property is what makes the prior and likelihood conjugate. For example, if you have a Gaussian (normal) prior distribution and you choose a Gaussian likelihood distribution, the resulting posterior distribution will also be a Gaussian distribution. This enables straightforward calculations and simplifies parameter estimation in Bayesian statistics.

The posterior will have the same distribution as A -- which is the definition of a conjugate prior.

Question 4

Linear regression by minimizing the sum of squared error is equivalent to maximizing the likelihood of data under a linear model with Gaussian noise.

ou Answered

- (selected) True
- False

True! We spent the second half of the lecture proving this.

Question 5

Write the following equation as a vector operation involving column vectors **x** and **y**:

$\sum_{i=1}^{d} x_i \, y_i$

Your Answer:

C = XY where X = [x1,x2,x3,...xd] (transpose of x) Y = [y1,y2,y3...,yd](transpose of y)

This is a dot product and could be written $< x, y >$ or $x^T y$

Question 6

Linear regression by minimizing the sum of squared errors (SSE) is robust to outliers.

True

ou Answered

False

False. As we showed in lecture, a single outlier point is sufficient to dramatically change the solution to ordinary least squares.

Question 7

How can our techniques for linear regression be used to fit non-linear functions of the input?

Your Answer:

We are able to play a little trick in order to use linear regression to fit non-linear functions. During linear regression, the transformation needs to be linear (So this means that our function for taking in inputs and giving an output needs to be of the form y = w1*i1 + w2*i1 + ... wn*in where w is the weight of the associated input, and i is the input value. This is linear because the transformation which occurs between the input and the output is linear.) But there is nothing which states that our inputs need to be linear. We are allowed to chose whatever we want to be as our input. So we can chose inputs which associate with a non-linear function.

We can apply basis functions to augment our data with non-linear transforms. This results in the data matrix gaining additional columns. Running standard linear regression on this new augmented data matrix works fine because the output is still a linear function of the now non-linear representation.

Question 8

What is regularization?

Your Answer:

Regularization is a technique used to prevent overfitting and improve a model's ability to work well with new data. Overfitting happens when a model gets too attached to the training data and can't handle new information. Regularization adds a penalty term to the model's loss function, encouraging it to have simpler and smoother parameter values. This makes the model more adaptable and better at making predictions.

Regularization is a way to express prior belief about what the parameters of a model should be. It is typically used to express a belief that the model should be as simple. There is a trade-off between the training error of a model and the simplicity of the model.

Question 9

In lecture, we fit the parameters of the logistic regression function by minimizing the _________________.

- log likelihood by setting the derivative to zero.
- log prior by using gradient ascent.
- log posterior by setting the derivative equal to zero.
- ou Answered: negative log likelihood using gradient descent.

negative log likelihood using gradient descent.

When we computed the gradient of the negative log likelihood for the logistic regression model, we were met with a system of non-linear equations that didn't let us write a closed-form solution. Instead, we used the gradient expression to perform gradient descent to minimize the negative log likelihood.

Question 10

When does $\sigma\left(w^T x\right)$ equal 0.5? And what does that imply about the decision boundary of the a logistic regression model?

Hint: σ is the logistic function.

Your Answer:

The logistic function is: σ(z) = 1 / (1 + e^(-z)) then we can plug in our value to the equation and solve for z :

1 / (1 + e^(-z)) = 0.5   ----------> z = 0

So, σ(w^Tx) equals 0.5 when z (w^Tx) is equal to 0.

Decision boundary:

It implies that the odds of data points belonging to either class are equal, this means that the model is uncertain and assigns equal probabilities to both classes at this boundary. This boundary represents the region where the model is indecisive and cannot confidently classify data points. Points on one side of the decision boundary are more likely to belong to one class, and points on the other side are more likely to belong to the opposite class. The decision boundary is a key concept in logistic regression, as it defines how the model separates data into different classes.

$\sigma\left(w^T x\right)$ equals 0.5 when $w^T x$ equals 0. This suggests that logistic regression has a decision boundary (the set of locations where both classes are equally likely) defined by the line where $w^T x = 0$.

Survey Score: **1** out of 1