DS 122 Practice Final Exam

Name:

BUID:

Problem A (3 points)

A company ships packages that are variable in weight, with an average weight of 15 lbs and a standard deviation of 10. Assuming that the packages come from a large number of customers so that it is reasonable to model their weights as independent random variables, find the probability distribution of the total weight of 100 packages.

Solution

Let X_i denote the weight of the i-th package and $S = X_1 + X_2 + \cdots + X_{100}$ be the total weight of 100 packages.
CLT: The distribution of S is approximately normal with mean $100 \cdot 15$ and variance $100 \cdot 10^2$. That is, $S \sim N(1500, 10000)$.

Problem B (8 Points)

You are running a (very small) experiment on coin flipping comparing a "control" coin to a "treated" coin. You flip each coin once and observe the first coin comes up heads and the second coin comes up tails. Let's call the probability the "control" coin comes up heads p_1 and the probability the treated coin comes up heads p_2.

You want to estimate the parameters p_1 and p_2 but you are short on time so you only consider the possibility that each coin is either fair ($p = 0.5$) rigged for tails ($p = 0.1$) or rigged for heads ($p = 0.9$). Starting with a uniform prior, what are the posterior probabilities for the two parameters for the three possible hypotheses?

Solution:

"Control" coin which came up heads:

Hypothesis	Prior	Likelihood	Unnormalized Posterior	Posterior
Coin is rigged for tails, $p_1 = 0.1$	1/3	0.1	1/30	1/15
Coin is fair, $p_1 = 0.5$	1/3	0.5	1/6	1/3
Coin is rigged for heads, $p_1 = 0.9$	1/3	0.9	9/30=3/10	3/5

"Treated" coin which came out tails:

Hypothesis	Prior	Likelihood	Unnormalized Posterior	Posterior
Coin is rigged for tails, $p_2 = 0.1$	1/3	0.9	9/30=3/10	3/5
Coin is fair, $p_2 = 0.5$	1/3	0.5	1/6	1/3
Coin is rigged for heads, $p_2 = 0.9$	1/3	0.1	1/30	1/15

Problem C (5 points)

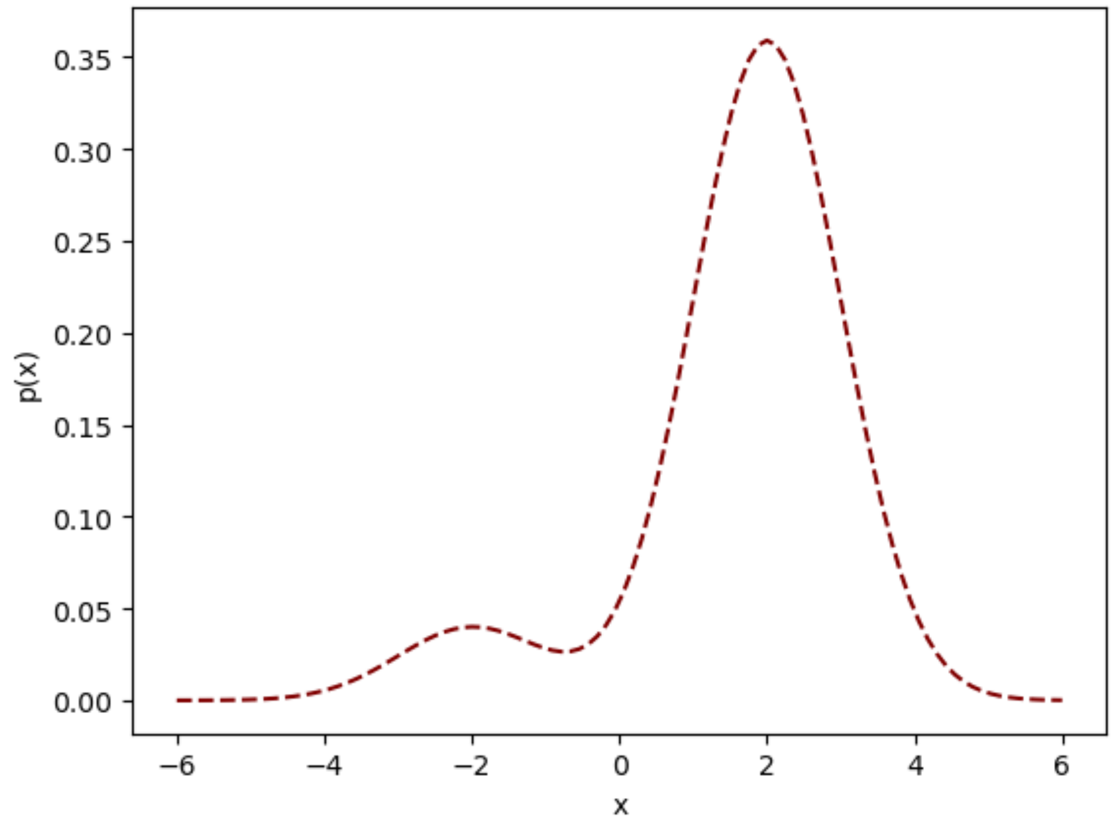
Last season the Boston Bruins scored 305 goals over the course of 82 games. So far this season the Bruins have scored 87 goals in 26 games (note: we treat all games as independent events). Using last years scoring rate as a prior modeled by a Gamma distribution, what is the posterior distribution of the the goal scoring rate for this season and what is the expected value of the distribution (hint: the expected value of a Gamma distribution is α/β)?

Solution:

Prior is $\theta \sim Gamma_{3.72,1}(\theta)$
Posterior is $\theta \sim Gamma_{90.72,27}(\theta)$
Expected value is 90.72/27=3.36. So we believe the Bruins goal scoring rate this season is 3.36 goals per game based on last years goal scoring rate and their perfomance so far this season.

Problem D (3 points)

You are using the Metropolis-Hastings algorithm to sample from the following complex distribution:



For simplicity, assume that in the plot above, the probability p(x) at x=0 is 0.05 and at x=3 is 0.3.

Assume candidates are selected using a normal distribution around the current state with a standard deviation of 1.

Consider a transition from to 3 to 0. What is the probability this transition will be accepted?

Solution:

Probabilty of acceptance of transition from j to i is given by:

$$a_{i,j} = min(1, \frac{\pi_i H_{j,i}}{\pi_j H_{i,j}})$$

Our transition matrix H is symmetric, so this reduces to:

$$a_{i,j} = min(1, \frac{\pi_i}{\pi_j}) = min(1, \frac{0.05}{0.3}) = 0.17$$

Problem E (7 points)

Let $x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}$ be a sample from a distribution with probability density

$$p_\theta(x) = \theta x^{\theta-1} \text{ for } x \in (0, 1)$$

and 0 otherwise. Here $\theta > 0$ is the unknown parameter. Determine the maximum likelihood

Solution

Let X represent the sample. Then the likelihood function is given by

$$p(X;\theta) = p(x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}; \theta) = \theta^4 \left(x^{(1)}\right)^{\theta-1} \left(x^{(2)}\right)^{\theta-1} \left(x^{(3)}\right)^{\theta-1} \left(x^{(4)}\right)^{\theta-1}.$$

The corresponding log-likelihood function is

$$\begin{aligned} \log p(X;\theta) &= \log\Big(\theta^4 \left(x^{(1)}\right)^{\theta-1} \left(x^{(2)}\right)^{\theta-1} \left(x^{(3)}\right)^{\theta-1} \left(x^{(4)}\right)^{\theta-1}\Big) \\ &= 4\log\theta + (\theta-1)\left(\log x^{(1)} + \log x^{(2)} + \log x^{(3)} + \log x^{(4)}\right) \\ &= 4\log\theta + (\theta-1)\sum_{i=1}^{4} \log x^{(i)}. \end{aligned}$$

To find the maximizer, we differentiate the log-likelihood function with respect to θ and set the result equal to zero:

$$\frac{\partial}{\partial\theta}\log p(X;\theta) = \frac{4}{\theta} + \sum_{i=1}^{4} \log x^{(i)} = 0.$$

Thus, $\hat{\theta} = -\frac{4}{\sum_{i=1}^{4} \log x^{(i)}}$.

Problem F (12 points)

With a primary diet of leaves, which are not very nutritional and are hard to digest, red howler monkeys spend most of their time eating and resting. Eating (E) and resting (R) are considered to be their primary states. Suppose that scientists observe that, on average, the monkeys spend 30% of their time eating and 70% resting. Suppose also that the biologists determined that if a monkey is eating at hour n, then there is a 60% probability that the animal will be eating the next hour, $n+1$. If a monkey is resting at hour n, then the next hour, $n+1$, it will remain resting with probability 80%.

Moreover, suppose that a group of scientists wants to study the behaviour of a red howler monkey in a more remote area. Knowing they will have limited opportunities of making visual observations, they attach a small microphone to one of the monkeys, whom they name Holly. The biologists discern that when hearing munching (M), the monkey is probably eating; but when hearing breathing (B) noises, the animal is likely to be at rest. The scientists analyze the noises once an hour. However, the microphone results are not always accurate. Besides background noises, such as from rain or other monkeys, a sleeping Holly might be moving her mouth. The scientists assume that there is 90% chance that if Holly is resting, then their system indicates breathing noises. They also assume that their system is only 80% accurate in detecting munching noises, while Holly is eating.

The first recording received by the scientists has been interpreted as munching. The recording that was received an hour later has been interpreted as breathing.

Part 1 (3 points)

Provide the initial state vector π, transition matrix A, and emission matrix B for the described problem.

Solution

$$\pi = \begin{bmatrix} 0.3 \\ 0.7 \end{bmatrix}, A = \begin{bmatrix} 0.6 & 0.2 \\ 0.4 & 0.8 \end{bmatrix}, B = \begin{bmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{bmatrix}.$$

Part 2 (7 points)

Perform the initialization and forward pass of the Viterbi algorithm.

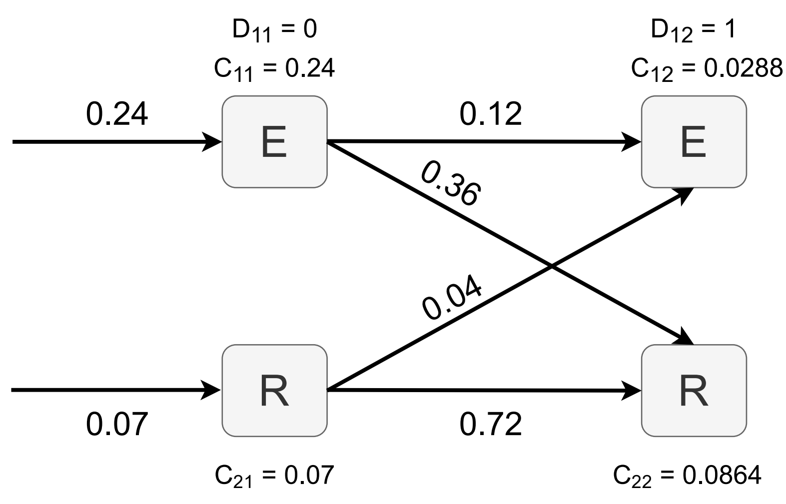
Solution

Initialization:

$$P(M|E)P(E) = 0.8 \cdot 0.3 = 0.24$$
$$P(M|R)P(R) = 0.1 \cdot 0.7 = 0.07$$
$$C = \begin{bmatrix} 0.24 & * \\ 0.07 & * \end{bmatrix}, D = \begin{bmatrix} 0 & * \\ 0 & * \end{bmatrix}.$$

Forward pass:

$$P(B|E)P(E|E) = 0.2 \cdot 0.6 = 0.12$$
$$P(B|R)P(R|E) = 0.9 \cdot 0.4 = 0.36$$
$$P(B|R)P(R|R) = 0.9 \cdot 0.8 = 0.72$$
$$P(B|E)P(E|R) = 0.2 \cdot 0.2 = 0.04$$
$$C_{12} = \max(0.12 \cdot 0.24, 0.04 \cdot 0.07) = 0.0288$$
$$C_{22} = \max(0.36 \cdot 0.24, 0.72 \cdot 0.07) = 0.0864$$



$$C = \begin{bmatrix} 0.24 & 0.0288 \\ 0.07 & 0.0864 \end{bmatrix}, D = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}.$$

Part 3 (2 points)

Find which primary states correspond to the first two recordings received by the scientists by performing the backward pass.

Solution

From the last column of C follows that $s = 2$. This leads to (R), Resting, being the second hidden state. Since $D_{22} = 1$, the first hidden state is E, Eating. Answer: (Eating, Resting)

Problem G (6 points)

You would like to find a slope β_1 and intercept β_0 for a line that minimizes the sum of squared error between predicted and observed values. Because your dataset has billions of datapoints, you decide to use sochastic gradient descent.

If you initilize the slope to -4 and the intercept to 0, set the learning rate to 0.1 and the first random data point you select is (2,5), what will be the new slope and intercept after one step of stochastic gradient descent?

Solution

The update for stochastic gradient descent for a single datapoint for the sum of squared error loss function where f is the model is given by:

$$w^{n+1} = w^m - \eta(f(x_i; w^n) - y_i)\nabla_w f(x_i, w^m)$$

In linear regression the model is $\hat{y}_i = x_i^T \beta$.

Putting that together we have:

$$\beta^{n+1} = \beta^n - \eta(x_i^T \beta^n - y_i)x_i$$

And using our selected data point and step size:

$$\beta^{n+1} = \begin{bmatrix} 0 \\ -4 \end{bmatrix} - 0.1(\begin{bmatrix} 1 \\ 2 \end{bmatrix}^T \begin{bmatrix} 0 \\ -4 \end{bmatrix} - 5)\begin{bmatrix} 1 \\ 2 \end{bmatrix}$$