CSCI 447/547: Homework 2

Due: 10/07/2021

Please submit your responses to the following questions on Moodle. You may work with a partner, and are encouraged to develop solution strategies together, but your actual submitted responses must be entirely your own. Using the internet is okay, but try to solve these on your own first, and be sure not to plagiarize! The objective of these exercises is not to get the correct answer, but rather to demonstrate that you understand the answer and the question to which you are responding. As such, please use complete sentences and develop your arguments fully. In particular, don't leave it to me to try to divine your meaning!

Problem 1: Full Bayesian Inference

We have seen that we can estimate the maximum likelihood estimators for a variety of distributions from some observations. However it is also possible to infer full posterior distributions, rather than just point estimators for some useful cases.

Coin flips

Consider first the problem of trying to infer the weighted-ness θ of a coin. In particular, use Bayes' theorem to show that the posterior distribution

$$P(\theta|X)$$

is Beta-distributed, assuming that the likelihood $P(X|\theta) = \prod_{i=1}^{m} \text{Bernoulli}(X_i;\theta)$ is independent and Bernoulli-distributed and the prior $P(\theta) = \text{Beta}(\theta;\alpha,\beta)$ is Beta-distributed with known hyperparameters α and β. (If you don't know where to start with this: first, look up the PDF for a beta distribution, which you should use for the prior. Multiply it with the likelihood function. Collect terms and perhaps rename these collections so that the resulting posterior again has the PDF of a beta distribution).

Normals (GRAD STUDENTS)

Use Bayes' theorem to show that the posterior distribution over the mean of a normal distribution

$$P(\mu|X)$$

is Normally-distributed, assuming that the likelihood $P(X|\mu,\sigma^2) = \prod_{i=1}^{m} \mathcal{N}(X_i;\mu,\sigma^2)$ is independent and Normally-distributed with known variance σ^2 and the prior $P(\mu) = \mathcal{N}(\mu;\mu_0,\sigma_0^2)$ is Normally-distributed with known hyperparameters μ_0 and σ_0^2. The approach is the same as above; however you'll need to use the technique "completing the square" to get the posterior in the form of a normal distribution.

Problem 2: Feature augmentation

(a)

Suppose that you have a dataset that looks like Figure 1, and you would to perform linear regression on it. Describe your approach for constructing a design matrix Φ for this problem. Specifically, what would you

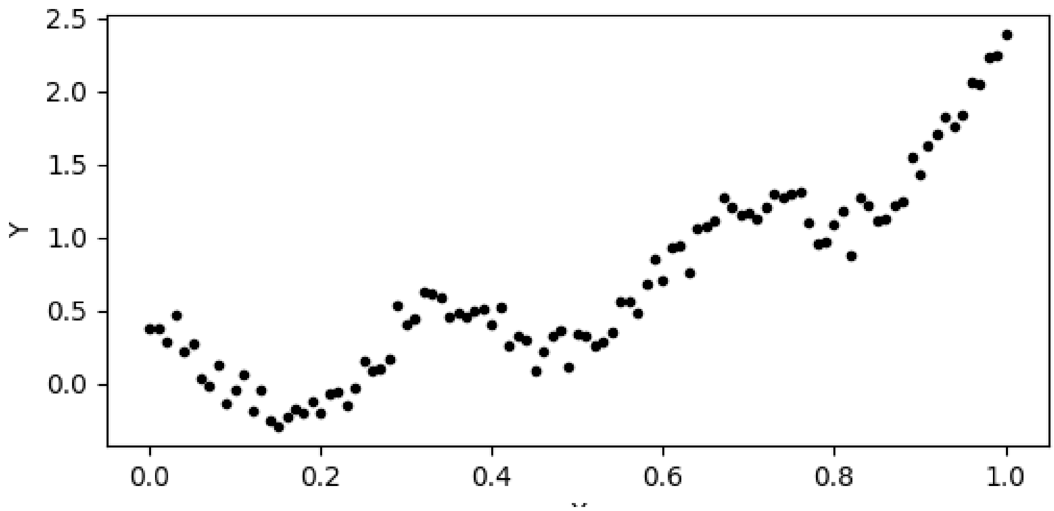

Figure 1: The dataset associated with Problem 2a.

place in each of the columns of Φ?

(b)

Same question as above. Would the same approach that you used in Part 1, also work here? Why or why not?

Problem 3: Gradient Descent

(a)

Consider the function

$$\mathcal{L}(\vec{\theta}) = \frac{1}{2}(\theta_1^2 - \theta_2)^2 + \frac{1}{2}(\theta_1 - 1)^2.$$

Derive an expression for the gradient of this function with respect to $\vec{\theta}$. Implement some code that uses gradient descent to find its minimum value. Create a plot similar to those we did in class that illustrates your algorithm's progress to this minimum beginning from at least 3 randomly selected initial positions.

(b) (GRAD STUDENTS)

Consider the function

$$\mathcal{L}(\theta) = -\theta_1 + \frac{1}{2}\sin(2\pi\theta_1)\sin(2\pi\theta_2).$$

Perform the same exercise as above. Does gradient descent always find the same solution for this function? Explain why or why not.

0.1 (c)

Describe, as specifically as you can, why it is not possible to solve directly for the minimum values of the functions that you worked with above. Describe how this relates to logistic regression.

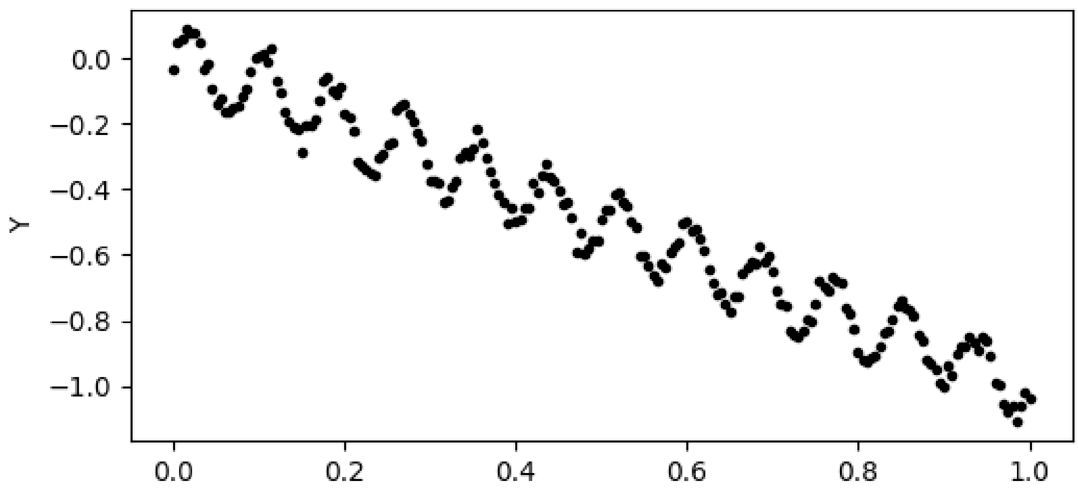

Figure 2: The dataset associated with Problem 2b.