Problem Set 5

CS 6375

Due: 5/5/2022 by 11:59pm

Note: all answers should be accompanied by explanations for full credit. Late homeworks will not be accepted.

Problem 1: Poisson Maximum Likelihood Estimation (30pts)

Consider a nonnegative, integer-valued random variable X that is distributed according to a Poisson distribution $X \sim \frac{\lambda^x e^{-\lambda}}{x!}$ for some real-valued parameter $\lambda > 0$.

1. Given data samples $x^{(1)}, \ldots, x^{(m)}$, what is the maximum likelihood estimate for λ?

2. Suppose now that you introduce a prior probability distribution, $\lambda \sim \frac{1}{5} \max\{-\lambda/10 + 1, 0\}$. What is the MAP estimate under this prior probability distribution?

3. Why might you not prefer a prior probability distribution like the above for this estimation task? What might be a better prior?

Problem 2: Logistic Regression (30pts)

For this problem, consider the Parkinson's data sets from homework 2.

1. Fit a logistic regression classifier to training data set. What is the accuracy on the test set? Explain why in standard logistic regression, without any type of regularization, the weights may not converge (even though the predicted label for each data point effectively does) if the input data is linearly separable.

2. Fit a logistic regression classifier with an ℓ_2 penalty on the weights to this data set using the validation set to select a good choice of the regularization constant. Report your selected constant, the learned weights and bias, and the accuracy on the test set.

3. Fit a logistic regression classifier with an ℓ_1 penalty on the weights to this data set using the validation set to select a good choice of the regularization constant. Report your selected constant, the learned weights and bias, and the accuracy on the test set.

4. Does ℓ_1 or ℓ_2 tend to produce sparser weight vectors?

Problem 3: I'm Thinking of a Number Between $-\infty$ and ∞(40pts)

Consider the classic guessing game in which, one person, the chooser, picks an arbitrary integer and another person, the guesser, wants to guess the number in as few attempts as possible. Each time the guesser makes a guess, the chooser says higher (if the chosen number is higher than the guess), lower (if the chosen number is lower than the guess), and correct otherwise. The game ends when the guesser guesses correctly.

1. A partially observed Markov decision process (POMDP) is a generalization of a Markov decision process in which the state of the environment is not directly observed by the agent. Instead, each time the agent takes an action a when the environment is in state s, the system transitions to a new state s' and the agent is given a reward r and an observation $o \in \Omega$, where Ω is the set of possible observations. A policy π in a POMDP is then a mapping from Ω to an action. Explain how to formulate the guessing game as a POMDP: you should provide the states, actions, observations, transition function, and reward function.

2. Is it possible to perform value iteration on your POMDP?

3. What do you think an optimal policy looks like for your POMDP?

4. Explain, in detail, how to apply deep Q-learning with a neural network to approximate the appropriate Q-value function for POMDPs. In particular, you should describe the neural network and how you can use it to find the "optimal" action, with respect to its Q-value, for a given observation.

Course Evaluation:

Please take a few minutes and go to `eval.utdallas.edu` and provide feedback on the course. In particular, your comments are used to guide future course content and course improvements.

Problem 1: Poisson Maximum Likelihood Estimation (30pts)

Consider a nonnegative, integer-valued random variable X that is distributed according to a Poisson distribution $X \sim \frac{\lambda^x e^{-\lambda}}{x!}$ for some real-valued parameter $\lambda > 0$.

1. Given data samples $x^{(1)}, \ldots, x^{(m)}$, what is the maximum likelihood estimate for λ?
2. Suppose now that you introduce a prior probability distribution, $\lambda \sim \frac{1}{5}\max\{-\lambda/10 + 1, 0\}$. What is the MAP estimate under this prior probability distribution?
3. Why might you not prefer a prior probability distribution like the above for this estimation task? What might be a better prior?

1. $f_X(x) = \frac{\lambda^x e^{-\lambda}}{x!}$

$\hat{\lambda}_{MLE} : L(\lambda; x_1 \cdots x_n) = \frac{\lambda^{x_1} e^{-\lambda}}{x_1!} \times \frac{\lambda^{x_2} e^{-\lambda}}{x_2!} \cdots \times \frac{\lambda^{x_n} e^{-\lambda}}{x_n!}$

$\ln(L(\lambda; x_1 \cdots x_n)) = \ln\left(\frac{\lambda^{x_1} e^{-\lambda}}{x_1!}\right) + \ln\left(\frac{\lambda^{x_2} e^{-\lambda}}{x_2!}\right) \cdots + \ln\left(\frac{\lambda^{x_n} e^{-\lambda}}{x_n!}\right)$

$= \sum_{i=1}^{n} \ln(\lambda^{x_i}) + \ln(e^{-\lambda}) - \ln(x_i!)$

$= \sum_{i=1}^{n} x_i \ln(\lambda) - \sum_{i=1}^{n} \lambda - \sum_{i=1}^{n} \ln(x_i!)$

$\ln(L) = \ln(\lambda) \sum_{i=1}^{n} x_i - n\lambda - \sum_{i=1}^{n} \ln(x_i!)$

$\frac{d}{d\lambda}\left(\ln(\lambda) \sum_{i=1}^{n} x_i - n\lambda - \sum_{i=1}^{n} \ln(x_i!)\right)$

$= \frac{1}{\lambda} \cdot \sum_{i=1}^{n} x_i - n - 0$

$= \frac{1}{\lambda} \cdot \sum_{i=1}^{n} x_i - n$

Let $\frac{1}{\lambda} \cdot \sum_{i=1}^{n} x_i - n = 0$

$\frac{1}{\lambda} \sum_{i=1}^{n} x_i = n$

$n\lambda = \sum_{i=1}^{n} x_i$

$\lambda = \frac{1}{n} \sum_{i=1}^{n} x_i$

2) $f_\lambda(\lambda) = \frac{1}{5} \max\left\{-\frac{\lambda}{10} + 1, 0\right\}$

$\lambda > 10, \ = 0$

$0 < \lambda < 10 \quad = -\frac{\lambda}{10} + 1$

$$f_\lambda(\lambda) = \begin{cases} 0 & \lambda \geq 10 \\ \frac{1}{5}\left(-\frac{\lambda}{10} + 1\right) & 0 < \lambda < 10 \end{cases}$$

$\hat{\lambda}_{MAP} = P(X|\lambda)\, P(\lambda)$

$\hat{\lambda}_{MAP} = P(x \mid \lambda) \left\langle \frac{1}{5} \max\left\{-\frac{\lambda}{10} + 1, 0\right\} \right\rangle$

$\hat{\lambda}_{MAP} = \left\langle \frac{1}{n} \sum_{n=1}^{n} x_i \right\rangle \left\langle \frac{1}{5} \max\left\{-\frac{\lambda}{10} + 1, 0\right\} \right\rangle$

$$= \begin{cases} 0 & \lambda \geq 10 \\ \frac{1}{5n}\left(\sum_{n=1}^{n} x_i\right)\left(\frac{-\lambda}{10} + 1\right) & 0 < \lambda \leq 10 \end{cases}$$

3)

By using this prior, any λ that is greater than or equal to 10 drives the entire estimation to 0, making the estimation meaningless. A better prior could be a conjugate pair of the Poisson distribution. The conjugate will yield a posterior distribution that still contains the same parameter as the original distribution.

task? What might be a better prior?

Problem 2: Logistic Regression (30pts)

For this problem, consider the Parkinson's data sets from homework 2. (Answered in Jupyter note book)

1. Fit a logistic regression classifier to training data set. What is the accuracy on the test set? Explain why in standard logistic regression, without any type of regularization, the weights may not converge (even though the predicted label for each data point effectively does) if the input data is linearly separable.
2. Fit a logistic regression classifier with an ℓ_2 penalty on the weights to this data set using the validation set to select a good choice of the regularization constant. Report your selected constant, the learned weights and bias, and the accuracy on the test set.
3. Fit a logistic regression classifier with an ℓ_1 penalty on the weights to this data set using the validation set to select a good choice of the regularization constant. Report your selected constant, the learned weights and bias, and the accuracy on the test set.
4. Does ℓ_1 or ℓ_2 tend to produce sparser weight vectors?

1. When performing log likelihood in the case where same sets of x yields different y, the slope never flats out. As a result, no maximum likelihood can be found, and MLE does not exist. Since the convergence of Logistic Regression require the existence of MLE, without so, convergence cannot be reached.

4.L1 produces sparser weight vectors, as it has higher variance.