COMP 652 - ECSE 608: Machine Learning - Assignment 3

Posted October 24, 2018
Due November 9, 2018

---

1. [30 points] **Structured bandits**

   This question explores the connections between the standard K-armed stochastic bandit setting and the linear stochastic bandit setting.

   (a) [10 points] Show that a K-armed stochastic bandit problem with Gaussian rewards can be formulated under the more general linear stochastic bandit setting. Explicit the structured action space $\mathcal{X}$ and the optimal parameter $\theta_\star$ in terms of disjoint action set $\mathcal{K}$ and expectations μ_k for $k \in \mathcal{K}$.

   (b) [10 points] Show how the regret for this linear setting connects to the regret in the finitely-armed setting.

   (c) [10 points] Show how the OFUL algorithm for this linear problem connects to the UCB algorithm in the finitely-armed setting.

---

2. [30 points] **Markov Random Fields**

   Consider the 2D spin glass model we discussed in class.

   (a) [10 points] Suppose that instead of connecting pixels in a 4-neighborhood, we want to connect them in an 8-neighborhood. Describe what the parameters of the undirected graphical model will be.

   (b) [10 points] Suppose that we want to use such a model to capture natural scenes in images. Describe the advantages and disadvantages of this model compared to connecting a pixel only to 4 neighbors.

   (c) [10 points] For the 2D Ising model connected as in class, write a Gibbs sampling algorithm, assuming that potentials are represented using linear energy functions and that evidence can be injected along the leftmost edge of the model. Assume the model is an $n \times n$ lattice.

---

3. [20 points] **Graphical Models**

   (a) A doctor gives a patient a (D)rug (drug or no drug) dependent on their (A)ge (old or young) and (G)ender (male or female). Whether or not the patient (R)ecovers (recovers or doesn't recover) depends on all D, A, G. In addition, A is independent of G.

   - Write down and draw the belief network (DAG model) for the above situation.
   - Explain how to compute p(recover | drug)
   - Explain how to compute p(recover | drug, young)

   (b) Consider the factorization of the joint distribution on three variables

   $$p(a, b, c) = p(a|b)p(b|c)p(c) \tag{1}$$

   where all the variables are binary. Draw the graphical model for this distribution. How many parameters are needed to specify this distribution? How many parameters do we need if we make no assumptions on the joint distribution?

---

4. [20 points] **A midterm preparation question**

   For each of the learning problems outlined below, specify what is the best learning algorithm to use and why. Note that you should give *one* algorithm for each problem, even if there are several correct answers.

(a) You have about 1000 training examples in a 6-dimensional continuous feature space and the output is binary. You only expect to be asked to classify 100 test examples.

(b) You have a problem similar to the previous one except that the inputs are now of dimension 10,000 and you know that the relation between inputs and outputs is linear.

(c) You want to predict the values of several stock indexes for the next 3 days from their values in the preceding week.

(d) You want to design an algorithm to predict whether a machine in a factory is likely to have a failure given various available measurements and you want to use this algorithm in the following way: if the likelihood of a failure is above 85%, you ask for a technician to do a preemptive maintenance check. You would also like to be able to leverage the fact that you know the likelihood of a random failure of any machine.