10-601 Introduction to Machine Learning

Machine Learning Department
School of Computer Science
Carnegie Mellon University

Oracles, Sampling, Generative vs. Discriminative

Matt Gormley
Lecture 19
March 20, 2018

Reminders

- **Midterm Exam**
  - Thursday Evening 6:30 – 9:00 (2.5 hours)
  - Room and seat assignments will be announced on Piazza
  - You may bring one 8.5 x 11 cheatsheet

Probabilistic Learning

Function Approximation

Previously, we assumed that our output was generated using a **deterministic target function:**

$$\mathbf{x}^{(i)} \sim p^*(\cdot)$$
$$y^{(i)} = c^*(\mathbf{x}^{(i)})$$

Our goal was to learn a hypothesis h(**x**) that best approximates c*(**x**)

Probabilistic Learning

Today, we assume that our output is **sampled** from a conditional **probability distribution:**

$$\mathbf{x}^{(i)} \sim p^*(\cdot)$$
$$y^{(i)} \sim p^*(\cdot|\mathbf{x}^{(i)})$$

Our goal is to learn a probability distribution p(y|**x**) that best approximates p*(y|**x**)

Robotic Farming

	Deterministic	**Probabilistic**
Classification (binary output)	Is this a picture of a wheat kernel?	Is this plant drought resistant?
Regression (continuous output)	How many wheat kernels are in this picture?	What will the yield of this plant be?

Oracles and Sampling

Whiteboard

- Sampling from common probability distributions
  - Bernoulli
  - Categorical
  - Uniform
  - Gaussian
- Pretending to be an Oracle (Regression)
  - Case 1: Deterministic outputs
  - Case 2: Probabilistic outputs
- Probabilistic Interpretation of Linear Regression
  - Adding Gaussian noise to linear function
  - Sampling from the noise model
- Pretending to be an Oracle (Classification)
  - Case 1: Deterministic labels
  - Case 2: Probabilistic outputs (Logistic Regression)
  - Case 3: Probabilistic outputs (Gaussian Naïve Bayes)

In-Class Exercise

1. With your neighbor, **write a function** which returns **samples from a Categorical**
   - Assume access to the `rand()` function
   - Function signature should be: `categorical_sample(theta)` where theta is the array of parameters
   - Make your implementation as **efficient** as possible!
2. What is the **expected runtime** of your function?

Generative vs. Discrminative

Whiteboard

- Generative vs. Discriminative Models
  - Chain rule of probability
  - Maximum (Conditional) Likelihood Estimation for Discriminative models
  - Maximum Likelihood Estimation for Generative models

Categorical Distribution

Whiteboard

- Categorical distribution details
  - Independent and Identically Distributed (i.i.d.)
  - Example: Dice Rolls

Takeaways

- One view of what ML is trying to accomplish is **function approximation**
- The principle of **maximum likelihood estimation** provides an alternate view of learning

- **Synthetic data** can help **debug** ML algorithms
- Probability distributions can be used to **model** real data that occurs in the world
  (don’t worry we’ll make our distributions more interesting soon!)

Learning Objectives

Oracles, Sampling, Generative vs. Discriminative

You should be able to…

1. Sample from common probability distributions
2. Write a generative story for a generative or discriminative classification or regression model
3. Pretend to be a data generating oracle
4. Provide a probabilistic interpretation of linear regression
5. Use the chain rule of probability to contrast generative vs. discriminative modeling
6. Define maximum likelihood estimation (MLE) and maximum conditional likelihood estimation (MCLE)