Nonparametric Bayes

Omar Gutiérrez

@trinogz

June 15, 2015

Mostly based on A Tutorial on Bayesian Nonparametric Models by Samuel J. Gershman.
Outline

Introduction

Example: clustering
  Traditional Approach
  Alternative Approach

Conclusions
Introduction

- What we do in ML is fitting a model to the data
- That is, we adjust the values of certain parameters
Figure 1: Linear Regression
Neural Networks

![Perceptron Diagram]

Figure 2: Perceptron

<table>
<thead>
<tr>
<th>$x_0$</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>...</th>
<th>$x_n$</th>
<th>$\sum f(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.5</td>
<td>0.1</td>
<td>0.1</td>
<td>$1 \times 0.5 + 1 \times -1 = -0.5$</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0.5</td>
<td>0.1</td>
<td>0.1</td>
<td>$1 \times 0.5 + 0 \times -1 = 0.5$</td>
</tr>
</tbody>
</table>
Hidden Markov Models

Figure 3: Hidden Markov Models
Bertrand Russell's Inductivist Turkey

Figure 4: A comparison of models
Bertrand Russell's Inductivist Turkey

Figure 5: A comparison of models
Bayesian Learning

\[ P(h|D) = \frac{P(D|h)P(h)}{P(D)} \] (1)
Maximum Likelihood Estimation

\[ h_{MAP} \equiv \arg \max_{h \in H} P(h|D) \]
\[ = \arg \max_{h \in H} \frac{P(D|h)P(h)}{P(D)} \]
\[ = \arg \max_{h \in H} P(D|h)P(h) \]
\[ h_{MLE} = \arg \max_{h \in H} P(D|h) \]
Data is a mess

- The articles in Wikipedia
- The species in the planet
- The hashtags on Twitter
How the problem is *sometimes* addressed

- Let’s start with the classic approach
- Let’s do clustering
- Let’s use Gaussian Mixture Models (GMM)
- We can fit several models and then compare them with some metric.
How the problem is *sometimes* addressed

(a) 2  
(b) 3

(c) 4  
(d) 5  
(e) 6

Figure 6: A comparison of clusterings classified with GMM
How the problem is *sometimes* addressed

**Figure 7**: Bayesian Information Criterion (BIC)
Another interesting approach is to use Bayesian Nonparametric (BNP) models.

BNP models will build a model than can adapt its complexity to the data.
Bayesian nonparametric models

Finite

Parametric

Bayesian models

Nonparametric

Infinite

Mixture

Chinese Restaurant Process

Latent factor

Indian Buffet Process

Dirichlet Processes!
Chinese Restaurant Process

- Infinite number of tables
- A sequence of customers entering the restaurant and sitting down
- The first customer enters and sits at the first table
- The second customer enters and sits...
  - at the first table with probability $\frac{1}{1+\alpha}$
  - at the second table with probability $\frac{\alpha}{1+\alpha}$
How we can *alternatively* approach the problem

Figure 8: Points classified with Infinite GMM
What else can be done?

Figure 9: Digit recognition (datamicroscopes)
What else can be done?

Figure 10: Topic Modeling (datamicroscopes)
Recap: Bayesian parametric vs nonparametric models

- **Traditional approach (finite)**
  - The number of parameters $\theta$ (e.g. clusters) is prespecified
  - We have a prior distribution over parameters $P(\theta)$
  - For example, in the Gaussian mixture model, each cluster will be modelled using a parametric model (e.g. Gaussian)

- **Bayesian nonparametric models**
  - We assume that there is an **infinite** number of latent clusters
  - A finite number of clusters is *inferred* from data
  - The number of clusters grow as new data points are observed
Libraries in Python

- Sklearn
- Datamicroscopes
What else to learn?

- What is the $\beta$ distribution?
- What is the Dirichlet distribution?
- Dirichlet process
References

- Machine Learning by Tom Mitchell
- A Tutorial on Bayesian Nonparametric Models by Samuel J. Gershman
- datamicroscopes library
Thank you

Questions?