Lecture Pattern Analysis

Part 01: Vocabulary, Probabilities, and Sampling

Christian Riess
IT Security Infrastructures Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg
April 17, 2021

Introduction

- We require some vocabulary for our communication about Machine Learning (ML)

- We also require a formal framework to discuss ML algorithms on a scientific basis
- Arguably the most widely used framework is probability theory

- We will also introduce our first algorithm, namely how to sample from a Probability Density Function (PDF)

Pattern Recognition Recap and Classification Vocabulary

- Remember the steps of the classical pattern recognition pipeline:

(Data) ⇢ Sampling → Prepro-cessing → Feature Extraction → Classi-fication ⇢ (Class)




$\rightarrow \mathbf{x}$ $\quad f(\mathbf{x})$ $\quad y$

- Fundamental ML assumption: good feature representations map similar objects to similar features
- Classifier training is virtually always **supervised**, i.e. a training sample is a tupel $(\mathbf{x_i}, y_i)$ (cf. lecture "Pattern Recognition")
- **Unsupervised** ML works without labels, i.e., it only operates on inputs $(\mathbf{x_i})$ Hence, unsupervised ML only works on the distribution of the features
- Geek info: fashionable variants are semi-supervised ML (some data has labels), self-supervised ML (auto-generate surrogate labels)

Recap on Probability Vocabulary

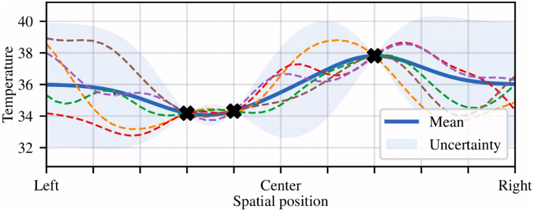
- We oftentimes operate with random variables X, Y
- Important vocabulary and equations are:

Joint distribution	$p(X, Y)$
Conditional distribution of X given Y	$p(X\|Y)$
Sum rule / marginalization over Y	$p(X) = \sum_Y p(X, Y)$
Product rule	$p(X, Y) = p(Y\|X) \cdot p(X)$
Bayes rule	$p(Y\|X) = \frac{p(X\|Y) \cdot p(Y)}{p(X)}$
Bayes rule in the language of ML	$\text{posterior} = \frac{\text{likelihood} \cdot \text{prior}}{\text{evidence}}$

- Please browse the book by Bishop, Sec. 1.2.3, to refresh your mind if necessary!

Sampling from a PDF

- Oftentimes, it is necessary to draw samples from a PDF
- Example:
  - Logistic Regression fits a single regression curve to the data (cf. PR)
  - Bayesian Logistic Regression fits a distribution of curves



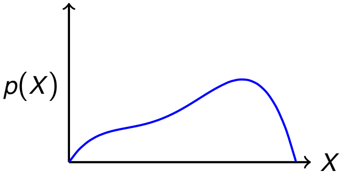
  The distribution is narrow at observations (crosses), and wider otherwise
  - Sample curves from the distribution to obtain its spread (“uncertainty”)
- Special PDFs like Gaussians have closed-form solutions for sampling
- We look now at a sampling method that works on **arbitrary PDFs**

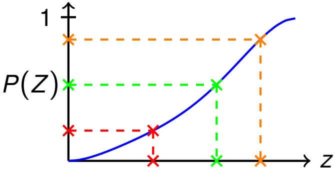
Idea of the Sampling Algorithm

- The key idea is to use the cumulative density function (CDF) $P(z)$ of $p(X)$,

$$P(z) = \int_{-\infty}^{z} p(X) \mathrm{d}X \tag{1}$$

- A sample uniformly drawn from the CDF y-axis intersects $P(z)$ at location z
- This z position is our random draw from $p(x)$:





Sampling Algorithm 如何通过均匀分布来采样服从指数分布的样本集

- Discretize the domain of the PDF $p(X)$
- Linearize $p(X)$ if it is multivariate
- Calculate the cumulative density function $P(z)$ of $p(X)$, the range of that CDF must be between 0 to 1
- Draw a uniformly distributed number u between 0 and 1
- The sample from the PDF is

u就是y轴的采样，满足均匀分布

$$z^* = \underset{z}{\operatorname{argmin}}\; u \cancel{>} P(z) \qquad (2)$$

找到使得u小于pz的最小z

<