*** Applied Machine Learning Fundamentals ***
Decision Theory

Daniel Wehner

SAP SE

August 21, 2019

Agenda August 21, 2019

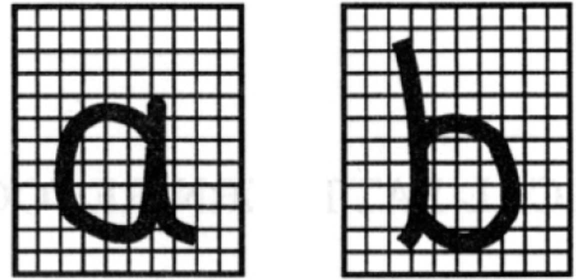
Section:

Bayesian Decision Theory

Statistical Methods

- Statistical methods assume that the process that 'generates' the data is governed by the **rules of probability**
- The data is understood to be a set of **random samples** from some underlying **probability distribution**
- The basic assumption about how the data is generated is always there, even if you don' t see a single probability distribution
- This is the reason for the name **statistical machine learning**

Running Example: Optical Character Recognition (OCR)



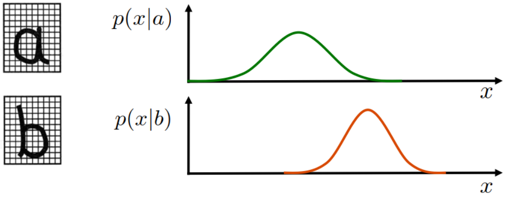
Goal: Classify a new letter so that the probability of a wrong classification is minimized

Class Conditional Probabilities

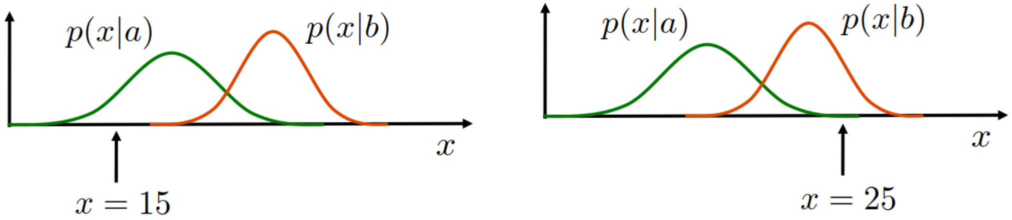
- First concept: **Class conditional probabilities**
- Probability of $\boldsymbol{x}$ given a specific class $\mathcal{C}_k$ is formally written as:

$$p(\boldsymbol{x}|\mathcal{C}_k) \in [0,1] \tag{1}$$

- $\boldsymbol{x} \in \mathbb{R}^m$ is a feature vector, e. g. # black pixels, height-width ratio, ...

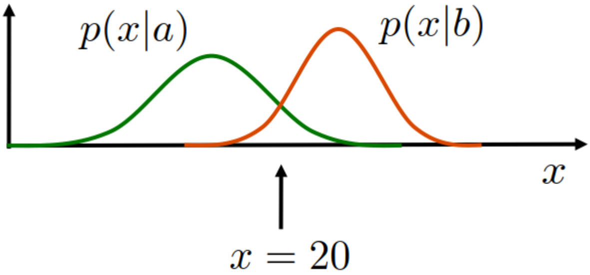


Class Conditional Probabilities (Ctd.)



If $\boldsymbol{x} = 15$ we would predict class a since $p(15|a) > p(15|b)$.
If $\boldsymbol{x} = 25$ we would output class b since $p(25|b) > p(25|a)$.

Class Conditional Probabilities (Ctd.)



We have a problem!

- **Which class should be chosen now?**
- The conditional probabilities are the same... ☠

Class Prior Probabilities

- Second concept: **Class priors**
- The prior probability of a data point belonging to a particular class $\mathcal{C}$

$$\mathcal{C}_1 \equiv a \qquad p(\mathcal{C}_1) = 0.75$$
$$\mathcal{C}_2 \equiv b \qquad p(\mathcal{C}_2) = 0.25$$

- By definition:
  - $0 \leqslant p(\mathcal{C}_k) \leqslant 1, \ \forall k$
  - The sum of all probabilities equals one: $\sum_{k=1}^{|\mathcal{C}|} p(\mathcal{C}_k) = 1$

How would you decide now?

- **The class prior is equivalent to a prior belief in the class label**

How to get the Prior Probabilities?

Count Count's advice:

Simply count the number of instances in each class!

Don't count **oranges**!

Bayesian Decision Theory: Bayes' Theorem

- What we actually want to compute: $P(\mathcal{C}_k|\mathbf{x}) \Rightarrow$ **Posterior probability**
- We can compute it by applying **Bayes' theorem**
- This is one of the **most important formulas (!!!)**

$$\overbrace{p(\mathcal{C}_k|\mathbf{x})}^{\text{Class posterior}} = \frac{\overbrace{p(\mathbf{x}|\mathcal{C}_k)}^{\text{Class cond.}} \cdot \overbrace{p(\mathcal{C}_k)}^{\text{Class prior}}}{\underbrace{p(\mathbf{x})}_{\text{Normalization term}}} = \frac{p(\mathbf{x}|\mathcal{C}_k) \cdot p(\mathcal{C}_k)}{\sum_{j=1}^{|\mathcal{C}|} p(\mathbf{x}|\mathcal{C}_j) \cdot p(\mathcal{C}_j)} \tag{2}$$

Section:

Naïve Bayes Classifier

Section:

Wrap-Up

Summary

Lecture Overview

Unit I: Machine Learning Introduction

Self-Test Questions

Recommended Literature and further Reading

Thank you very much for the attention!

Topic: *** Applied Machine Learning Fundamentals *** Decision Theory
Date: August 21, 2019

Contact:

Daniel Wehner (D062271)
SAP SE
daniel.wehner@sap.com

Do you have any questions?