(Session- 2018-19)

Introduction to Bayesian Learning

Introduction

- Allows us to combine observed data and prior knowledge
- Provides practical learning algorithms
- It is a generative (model based) approach, which offers a useful conceptual framework
  - This means that any kind of objects (e.g. time series, trees, etc.) can be classified, based on a probabilistic model specification

Bayes' Rule

$$p(h \mid d) = \frac{P(d \mid h)P(h)}{P(d)}$$

Understanding Bayes' rule
d = data
h = hypothesis
Proof. Just rearrange:
$p(h \mid d)P(d) = P(d \mid h)P(h)$
$P(d, h) = P(d, h)$
the same joint probability
on both sides

Who is who in Bayes' rule

$P(h)$: prior belief (probability of hypothesis *h* before seeing any data)
$P(d \mid h)$: likelihood (probability of the data if the hypothesis *h* is true)
$P(d) = \sum_h P(d \mid h)P(h)$: data evidence (marginal probability of the data)
$P(h \mid d)$: posterior (probability of hypothesis *h* after having seen the data *d*)

Does patient have cancer or not?

- A patient takes a lab test and the result comes back positive. It is known that the test returns a correct positive result in only 98% of the cases and a correct negative result in only 97% of the cases. Furthermore, only 0.008 of the entire population has this disease.
  - What is the probability that this patient has cancer?
  - What is the probability that he does not have cancer?
  - What is the diagnosis?

$$P(cancer) = .008,\ P(\neg cancer) = .992$$
$$P(+ \mid cancer) = .98,\ P(- \mid cancer) = .02$$
$$P(+ \mid \neg cancer) = .03,\ P(- \mid \neg cancer) = .97$$
$$P(cancer \mid +) = \frac{P(+ \mid cancer)P(cancer)}{P(+)}$$
$$P(\neg cancer \mid +) = \frac{P(+ \mid \neg cancer)P(\neg cancer)}{P(+)}$$

Maximum A Posteriori (MAP) Hypothesis

$$P(h \mid D) = \frac{P(D \mid h)P(h)}{P(D)}$$

The Goal of Bayesian Learning: the most probable hypothesis given the training data (Maximum A Posteriori hypothesis)

$$\begin{aligned} h_{MAP} &= \arg\max_{h \in H} P(h \mid D) \\ &= \arg\max_{h \in H} \frac{P(D \mid h)P(h)}{P(D)} \\ &= \arg\max_{h \in H} P(D \mid h)P(h) \end{aligned}$$

Thank you…