

Tutorial 3: Bayesian reasoning; classification

Computing probabilities from a joint probability distribution

1. **Quiz question:** Consult the joint distribution of *Toothache*, *Cavity* and *Catch* from Lectures 4 and 5:

	<i>toothache</i>		\neg <i>toothache</i>	
	<i>catch</i>	\neg <i>catch</i>	<i>catch</i>	\neg <i>catch</i>
<i>cavity</i>	.108	.012	.072	.008
\neg <i>cavity</i>	.016	.064	.144	.576

Calculate the following:

- (a) $P(\text{Catch} = \text{false} | \text{Cavity} = \text{true})$.
- (b) Probability distribution $p(\text{Cavity})$.
- (c) Probability distribution $p(\text{Toothache} | \text{Cavity} = \text{true})$.

Independence and Bayes' rule

2. What does it mean to say that two random variables in a probability model are **independent**?
3. Why is it useful to find random variables which are independent in a joint probability distribution?
4. What is **conditional independence**, and why is it useful?
5. What is **Bayes' rule**, and why is it useful?

Naive Bayes' Classifier

6. Say we are building a spam filter, in which we identify a set of **spam words** which are common in spam, and a set of **ham words** which are common in legitimate

(ham) email. The random variables in this situation are $Spam$ ($=spam$ or ham), and $W_1 \dots W_n$ (each of which is a Boolean variable signalling the presence/absence of each spam/ham word). We have a corpus of n known spam emails and one of m known ham emails. The probabilities we need to implement our filter are $p(w_i|spam)$ and $p(w_i|ham)$, as well as $p(spam)$ by itself.

- (a) How could we gather the probabilities we need from a corpus of known spam and ham emails?
- (b) How could we use Bayes' rule to derive the probability of a new document being spam given the presence/absence of the chosen spam/ham words—i.e. $\mathbf{P}(Spam|W_1 \wedge \dots \wedge W_n)$?
- (c) Assume we want to build a naive Bayes model of the email data.
 - i. What conditional independence assumption will we be making?
 - ii. Does this assumption seem reasonable?
 - iii. What simplification do these assumptions allow to the statement derived in part (b) of this question?

Decision Tree Classifier

7. **Quiz:** Consider the following set of example data, concerning a fictitious South Island Super 12 rugby team.

Play_at_home	Pitch_muddy	Opposition_from_S_Africa	Win
T	T	F	T
T	F	F	F
T	F	T	T
T	T	T	T
T	F	T	T
F	T	T	F
F	F	F	F
F	T	T	F
F	F	F	F

We are trying to learn a function which takes as input the Boolean variables Play_at_home, Pitch_muddy and Opposition_from_S_Africa, and which generates as output the variable Win.

- (a) Write down a decision tree which is consistent with these examples.
- (b) Explain the principle which determines how an *optimal* decision tree is built. (Just informally: you don't have to explain the maths.)