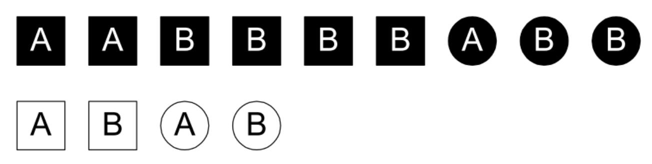
Tutorial: Probability Models[1]

Machine Learning

February 8, 2017

[1]Questions from various sources including Neapolitan and Jiang, "Contemporary AI", CRC Press (2012)

1. You are given a set of 13 squares and circles, 9 of which are coloured black and the rest are coloured white.
   Each object also has either the letter “A” or “B” on it. There are: 2 black squares with an A, 4 black squares with a B and 1 black circle with an A. Of the remaining, there is 1 white square and 1 white circle each with an A. Here is a diagrammatic representation:



   Let *Black* denote the set of black objects, *White* denote the set of white objects, *Square* denote the set of square objects, *A* the set of objects with an “A” and so on. Assuming all objects are equally likely (the so-called Principle of Indifference):

(a) What is $P(A)$?
(b) What is $P(A|Square)$?
(c) Are A and $Square$ independent?
(d) Are A and $Black$ independent?
(e) Are A and $Square$ conditionally independent given $Black$?
(f) Are A and $Square$ conditionally independent given $White$?
(g) The Law of Total Probability gives us: $P(A) = P(A, White) + P(A, Black)$. Verify that the law holds in this case.
(h) Using a probability-tree, calculate $P(Black|A)$.
(i) Using Bayes' Rule, calculate $P(Black|A)$

Basic Probability III

2. There are two urns (Urn1 and Urn2). Urn1 has 2 red marbles and 2 blue marbles. Urn2 has 1 red and 3 blue marbles.[2] The urn labels are now covered and a coin is flipped to select an urn. Having selected an urn, we draw a marble from the urn. The marble is red. What is the probability that the urn selected was Urn1?

3. In a typical English summer, the probability that the temperature falls below 10 degrees Celsius is 0.4. In that case, the English cricket team wins with probability 0.75. The probability that the temperature is between 10 and 30 degrees Celsius is 0.4, in which case the English team wins with probability 0.65. The probability that the temperature is greater than 30 degrees is 0.2 and in that case, the English team wins with probability 0.55. You have just received an

Basic Probability IV

SMS saying the English team has won. What is the probability that the temperature was below 10 degrees?

[2] Some of these exercises are from M. Cargal, "Discrete Mathematics for Neophytes".

Probability Distributions I

4. The probability mass function of a discrete r.v. is as follows:

$$p(X = x) = \begin{cases} 1/3 & x = -1, 0, 1 \\ 0 & \text{otherwise} \end{cases}$$

   What is $\mu_X = E(X)$?

5. You are told $Var(X) = E[(X - \mu_X)^2]$. What is $Var(X)$ for the r.v. in the above?

6. Repeat the calculations for the following mass function:

$$p(X = x) = \begin{cases} 1/3 & x = -2, 0, 2 \\ 0 & \text{otherwise} \end{cases}$$

   Why does the variance increase?

Probability Distributions II

7. Let X be the random variable denoting the number of dots that come up on the throw of a six-sided die. What is $E(X)$? (Are store-bought dice uniform?)

8. Let X be a random variable denoting the number of successes in n i.i.d. Bernoulli trials, each with probability p of success. What is $E(X)$?

9. Let X be an exponential random variable with pdf $f(X = x) = \lambda e^{-\lambda x}$ $(x > 0)$. What is $E(X)$? What is $E(X^2)$? Recall: integration by parts:

$$\int u dv \; = \; uv - \int v du$$

Probability Distributions III

10. A continuous real-valued variable has a power-law p.d.f. if $p(x) = Cx^{-\alpha}$ $(\alpha > 0)$. In fact, this function diverges as $x \to 0$: so how can it be a p.d.f. ?

11. Find an expression for C in the (modified) p.d.f. in the previous question.

12. Find the expected value for the random variable having the (modified) power-law p.d.f.

13. Power-laws with $\alpha \leq 2$ have no finite mean. This means that as we start taking more and more samples from such populations, we will start to see the mean diverge. How can this happen?

14. Similarly show that for $\alpha \leq 3$, there is no finite variance.

Maximum Likelihood I

15. You have a sample of n observations $x_1, x_2, \ldots, x_n$ from data that appear to fit a binomial distribution
with parameters N and p. Assuming N is known, derive the maximum likelihood estimate for p in terms of N, n, and the x_i.

16. Let $x_1, x_2, \ldots, x_n$ be a sample of observations from a Poisson distribution with parameter λ. Find the maximum likelihood estimate of λ in terms of the x_i and n.

17. Let $x_1, x_2, \ldots, x_n$ be a sample from an exponential distribution, which has a density function $f(X = x) = \lambda e^{-\lambda x}$ $(x > 0)$. Derive a maximum likelihood estimate of λ in terms of the x_i and n.

Maximum Likelihood II

18. Let $x_1, x_2, \ldots, x_n$ be observations from a normal distribution with parameters μ and σ^2. Derive maximum likelihood estimates of μ and σ^2.

Logistic Regression I

Simple limear regression deals with the problem of fitting a line $Y = a + bX$ for a set of points $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$. The least-squares estimates of b and a are:

$$b = \sum (x_i y_i) / \sum {x_i}^2$$

where $x_i = (X_i - \overline{X})$ and $y_i = (Y_i - \overline{Y})$; and

$$a = \overline{Y} - b\overline{X}$$

This extends naturally to *weighted* simple linear regression, in which each point has a weight w_i. The least-square estimates of b and a are then:

$$b = \sum w_i (x_i y_i) / \sum w_i {x_i}^2$$

Logistic Regression II

where $x_i = (X_i - \overline{X}_w)$ and $y_i = (Y_i - \overline{Y}_w)$; and

$$a = \overline{Y_w} - b\overline{X_w}$$

The means are now weighted averages:

$$\overline{X}_w = \frac{\sum w_i x_i}{\sum w_i} \quad \overline{Y}_w = \frac{\sum w_i y_i}{\sum w_i}$$

Clearly, if the $w_i = 1$, the ordinary linear regression results.

19. We will use weighted linear regression to build a linear model for the log-odds of Y when Y takes on one of two values: 0 and 1. For any value of $X = X_i$, the *odds* of Y (actually the odds of $Y|X = X_i$) is the ratio $P(Y = 1|X = X_i)/P(Y = 0|X = X_i)$. It is therefore simply the ratio of the number of $Y = 1$ entries for $X = X_i$ to the

Logistic Regression III

number of $Y = 0$ entries for $X = X_i$. This procedure is *simple logistic regression*.

Here is a partially completed table about a dataset:

i	ii	iii	iv	v	vi	vii
X	Y		Total	$P(Y=1\|X)$	$Odds(Y)$	$LogOdds(Y)$
	0	1				
28	4	2				
29	3	2				
30	2	7				
31	2	7				
32	4	16				
33	1	14				

(a) Complete the table. (b) Using the total for each X_i, Y_i as the weight w_i, obtain the weighted linear regression line $LogOdds(Y) = a + bX$. (c) What is the predicted probability for $X = 31$?

Simple Bayes I

20. The following table represents data collected by some machine-learning researchers at Wimbledon.

Day	Outlook	Temperature	Humidity	Wind	Play
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

Simple Bayes II

From Bayes' Rule (with some simplification of notation):

$$
\begin{aligned}
P(Yes|Sunny, Cool, High, Strong) &= \frac{P(Yes)P(Sunny, Cool, High, Strong|Yes)}{P(Sunny, Cool, High, Strong)} \\
&\propto P(Yes)P(Sunny, Cool, High, Strong|Yes) \\
P(No|Sunny, Cool, High, Strong) &= \frac{P(No)P(Sunny, Cool, High, Strong|No)}{P(Sunny, Cool, High, Strong)} \\
&\propto P(No)P(Sunny, Cool, High, Strong|No)
\end{aligned}
$$

Assume that the attributes Outlook, Temperature, Humidity and Wind are conditionally independent of each other given the value of the target attribute Play.
Using the data recorded, estimate the probability of play on Day 15, which has the following forecast:

$$\langle Outlook = Sunny, Temperature = Cool, Humidity = High, Wind = Strong\rangle$$