Machine Learning – Tutorial 1

Model Answers

1. The frequentist answer:

Pick the value of p that makes the observation of 53 heads and 47 tails most probable:

$$
\begin{aligned}
P(D) &= p^{53}(1-p)^{47} \\
\frac{dP(D)}{dp} &= 53p^{52}(1-p)^{47} - 47p^{53}(1-p)^{46} \\
&= \left(\frac{53}{p} - \frac{47}{1-p}\right)\left[p^{53}(1-p)^{47}\right] \\
&= 0 \;\; if \;\; p = .53
\end{aligned}
$$

If coin is tossed once, is p=1 a sensible answer?

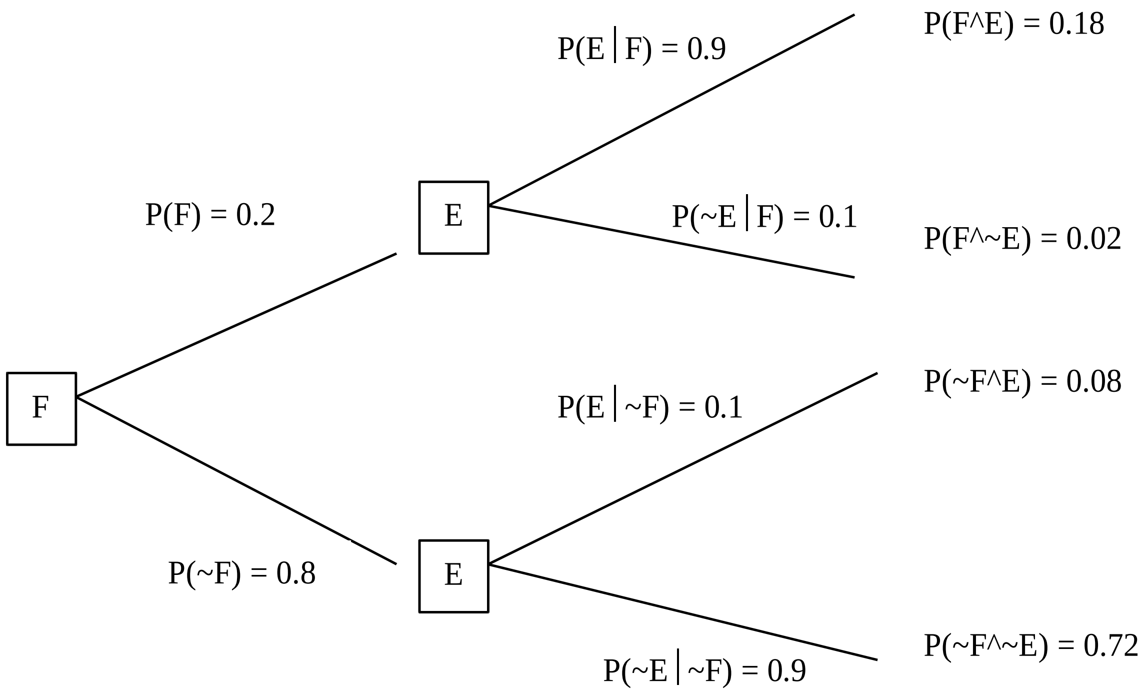
If we don’t have much data, we are unsure about p. Our computations of probabilities will work much better if we take this uncertainty into account: start with a prior distribution over p.

In this case we use a uniform distribution (remember: we know nothing about coins!). Then, multiply the prior probability of each parameter value by the probability of observing *heads* given that value. Then scale up all of the probability densities so that their integral comes to 1. This gives the posterior distribution.

After 53 heads and 47 tails we get a very sensible posterior distribution that has its peak at 0.53 (assuming a uniform prior).

2.

E = the event that a driver tests positive
F = the event that a driver has smoked cannabis in the last 72 hrs



From Bayes' theorem:

$$P(F\mid E)=\frac{P(E\mid F)*P(F)}{(P(E\mid F)*P(F))+(P(E\mid\sim F)*P(\sim F))}$$

$$P(F\mid E)=\frac{0.9*0.2}{(0.9*0.2)+(0.1*0.8)}=\frac{0.18}{0.18+0.08}=\frac{0.18}{0.26}=0.69$$

3.

B = voter from Bury, C = Croydon, D = Dover, W = voted for winner

P(D|W) = P(D^W)/P(W)

P(D^W) = 0.51 * 0.16 = 0.0816

P(W) = P(B^W)+P(C^W)+P(D^W) =

P(B)P(W|B)+P(C)P(W|C)+P(D)P(W|D) =

0.46*0.61+0.38*0.88+0.16*0.51= 0.6966

P(D|W) = 0.0816/0.6966 = 0.1171

4.

Let's put this in notation:
A = "studied AI"
B = "studied CS"
C = "didn't study AI or CS"
D = "knows what an NN is"

The question is to find the chance that George knows what a neural network is:
$p(D) = ?$

80% of the people who studied AI know what a neural network is, so the probability that someone knows what an NN is given that they studied AI is 0.8:
$p(D|A) = 0.8$

40% of the people who studied CS know what an NN is:
$p(D|B) = 0.4$

10% of the rest know what an NN is:
$p(D|C) = 0.1$

Now, I think George studied AI with 50% probability, and CS with 20%:
$p(A) = 0.5$
$p(B) = 0.2$

A, B and C are exhaustive and mutually exclusive (George studied only one of these or none), so we can deduce that:
$p(C) = 1 - p(A) - p(B) = 1 - 0.5 - 0.2 = 0.3$

We want to find p(D). Applying Bayes' theorem:

$p(D \wedge A) = p(D|A)\, p(A) = 0.8 \times 0.5 = 0.4$
$p(D \wedge B) = p(D|B)\, p(B) = 0.4 \times 0.2 = 0.08$
$p(D \wedge C) = p(D|C)\, p(C) = 0.1 \times 0.3 = 0.03$

Since A, B and C are exhaustive and mutually exclusive:

$p(D) = p(D \wedge A) + p(D \wedge B) + p(D \wedge C) = 0.4 + 0.08 + 0.03 = 0.51.$

5.

The facts are:
A = "picked the fake coin"
B = "the coin fell in 'head' 10 times in a row"

I want to know the probability that the coin that fell heads 10 times is the fake one; that is, the probability that a coin is the fake one given that it has fallen heads 10 times:

$p(A|B) = ?$

The probability of picking the fake coin is 1 in 1000:
$p(A) = 1/1000$
... so the probability of picking a normal coin is 999 in 1000:
$p(\sim A) = 999/1000$

The probability of a fake coin falling heads 10 times in a row is 1:
$p(B|A) = 1$

The probability of a normal coin falling heads 10 times in a row is 1/2^10:
$p(B|\sim A) = 1/2^{10} = 1/1024$

We want to find out p(A|B). Bayes' theory says:
$p(A|B)\ p(B) = p(A \wedge B)$
$p(A|B) = p(A \wedge B) / p(B)$ (1)

Bayes' theory (again) says that $p(A \wedge B) = p(B|A)\ p(A) = 1 \times 1/1000 = 1/1000$
...and also that $p(\sim A \wedge B) = p(B|\sim A)\ p(\sim A) = 1/1024 \times 999/1000 = 0.976/1000$, which I'll approximate to 1/1000.

Finally, from the lecture:
$p(B) = p(A \wedge B) + p(\sim A \wedge B)$
$p(B) = 1/1000 + 1/1000 = 2/1000$
which we replace in equation (1):

$p(A|B) = 1/1000\ /\ (2/1000) = 1/2.$
So there is a probability of (approx.) 50% that the coin we picked is the fake.

6.

Your possible actions are: turn witness and if John also does you get 3 years, if he doesn't you get zero. Or you may refuse to turn witness and if John does you get 6 years, and if he doesn't you get 1 year.

According to classical decision theory, you should choose the action that has highest expected utility, in this case the action that minimizes the number of years you expect to spend in jail.

For example, because John is your friend, you might believe:

John turns witness with probability 10%.

John refuses to turn witness with probability 90%.

The best option is for you both to refuse to turn witness against the other. However, you turning witness is your less risky option.

Suppose the probability of John turning witness is P. So, if you turn witness and John also does you get 3 years with probability P (3P); if he doesn't you get 0 years (0(1-P)), for a total of 3P.

If you don't turn witness but John does you get 6P, otherwise 1(1-P), for a total of 5P+1. Since 3P is less than 5P+1 for any P, you're always better off turning witness against him (morality aside).

Unfortunately, if he has also studied Decision Theory he will also turn witness against you, and you'll both get 3 years.