1 Supervised learning

Supervised learning has a goal of mapping inputs x to outputs $yE1....C$ where C is the number of classes. If C = 2 it is binary classification, if C ¿ 2 it is multi class classification.

One way to formalise the idea is function approximation. We are mapping y = f(x) where the goal is to learn f from some set of data.

2 Probability

3 Basic probability

4 Basics

Union means that "A or B" can occur. Formally $P(AnB) = e$ where e is the empty set. If events are not mutually exclusive, the union $P(AorB) = P(A) + P(B) - P(AnB)$ (taking off the intersection because otherwise we are counting the probability twice) If the events are mutually exclusive, the union is $P(A) + P(B)$

If two events are mutually exclusive $P(AnB) = 0$

We say that X and Y are unconditionally independent if P(X, Y) = P(X)P(Y)

Joint probabilities

$P(A, B) = P(AnB) = P(A|B)P(B)$ Specifically, P(A, B) is a probability distribution and P(A, B) is a set of probabilities. This is sometimes known as the *productrule*.

We can define the marginal distribution as follows (only if they are conditionally dependent): $\sum P(A, B) = \sum P(A|B = b)P(B = b)$ If P(A n B) = P(A)P(B) then we can say the probabilities are independent.

Conditional probability

Event A occurring, given than B has occurred. $P(A|B) = \frac{P(A,B)}{P(B)}$

Bayes

Combining conditional probability with joint probability yields bayes theorem: $P(X = x|Y = y) = \frac{P(X=x),Y=y}{P(Y=y)} = \frac{P(X=x)P(Y=y|X=x)}{\sum P(X=x`)P(Y=y|X=X`)}$

5 Distributions

5.1 Binomial and Bernoulli distribution

The binomial distribution is discrete and only takes values yes/no with the probabilities and number of trials. It takes the values k, n and p where k is the number of successes in n trials where each success has probability k.

X B(n, p) P(X = r) = (n r) $p^r q^(n - r)$

Bernoulli distribution is a discrete distribution that takes binary values, '1' with probability p and '0' with probability '1-p'. It is a special case of the binomial distribution, where n = 1.

5.2 Beta-binomial distribution

Beta-binomial distribution is a family of discrete probability distributions that arises when the probability of success in a bernoulli distribution is either fixed or known and the probabilities of success at each trial is not fixed(as it is in Bernoulli) but instead follows a beta-distribution.

The beta-binomial is a one-dimensional version of the dirichlet-multinomial distribution. The multivariate generalisation of the beta-distribution is the dirichlet distribution.

6 TODO

- Separate probability as conditional, independent, dependent, etc.

7 Definitions

Definition: Binary classification

When there are only two classes(C = 2) in which we are classifying data

Definition: Multiclass classification

When there are more than two classes we can classify data to. C ¿ 2

Definition: Multi-label classification

When labels are not mutually exclusive, data could be classified to two classes for example "strong" and "tall"

Definition: Mutually exclusive

Events cannot occur at the same time, for example tossing a coin obtains either Heads or Tails but it cannot be both

Definition: Independent events

Events are independent if the occurrence of one events gives us no information about whether or not the other event will occur. Events have no influence on each other.

Definition: Conditionally independent

If event C has not occurred, P(A—B) = P(A) and P(B—A) = P(B). Eventually the event C occurs, and now if the event A occurs, the probability of B occuring will decrease and vice versa. The two probabilities become conditionally dependent because their probability of occurrence is dependent on the other's events