1 Basis expansion

1.1 Polynomial regression

To extend linear regression to settings where the function is non-linear, basis functions are used to construct a larger space of functions. A basis function (or number of basis functions) are defined on the input X. For example

$$h_1(X) = 1, h_2(X) = X, h_3(X) = X^2 \tag{1}$$

forms three basis functions which is used to transform the input. This transformation replaces the standard linear model

$$f(X) = \theta_0 + \theta_1 X \tag{2}$$

with new inputs

$$[h_1(X), h_2(X), h_3(X)]^T \to [1, X, X^2]^T \tag{3}$$

to get a polynomial function

$$\begin{aligned} f(X) &= \theta_0 h_1(X) + \theta_1 h_2(X) + \theta_2 h_3(X) && (4) \\ &= \theta_0 + \theta_1 X + \theta_2 X^2 && (5) \end{aligned}$$

Now the same linear regression can be run on the new input as the output is a linear combination of functions, which a linear regression model can still be applied on. The functions may be non-linear, but the input space has been transformed through the non-linear mapping. Of course, this can naturally be extended to m polynomials, but generally using more than 3 or 4 polynomials leads to high chances of overfitting.

1.2 Piecewise linear regression

Basis expansions are in fact more general than polynomial expansion. Basis functions can also be defined to create a piecewise linear regression model.

1.3 Regression splines

2 Bayesian classification

In Bayesian terms, input variables are called **evidence**. The Bayes rule allows the probability and evidence to be expressed into simpler distributions:

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \tag{6}$$

where

- $P(C_i|X)$ is the **posterior probability**. In other words, if we know X happened, how likely is C?
- $P(X|C_i)$ is the **likelihood**. If we know that the right answer is C, how likely is observation X?
- $P(C_i)$ is the **prior**, which is how often C_i is the correct answer in total. This can often be measured or estimated.
- $P(X)$ is the **probability of evidence**, which is how often this particular observation is obtained.

Many classifiers attempt to calculate the posterior directly, for example logistic regression. The underlying assumption is that the priors will not change.

2.1 Maximum likelihood classifier

The maximum likelihood classifier picks the class most likely to generate X. It only makes sense when the classes are approximately equally likely.

$$\hat{Y} = \text{argmax}_c P(X|C) \tag{7}$$