**ST3009 – Statistics**

**Linear Regression**

* **Line of Best Fit Guess:**
* **Parameters:**(Y-intercept), (Slope)
* **Cost Function:** The cost function represents the distance of each actual data point () from the line of best fit (.
* **Goal:** Select a y-intercept () and a slope () for our line of best fit that minimises i.e gives us the most accurate line of best fit.
* **Adding Noise:** Most of the time our sample data is affected by randomly distributed Gaussian noise M. This can be factored into our line of best fit simply by:

Our training data can now be considered as:

* **Gaussian RV:** A Gaussian random variable Z with mean and variance has pdf:

Take for example an M with mean 0 and variance 1, we can then assume:

* **Bayes Rule:** Using Bayes rule we can infer the posterior, likelihood and prior:

Posterior: The probability of parameter given data

Likelihood: The probability of data given parameter

Prior: Probability of prior

* **Likelihood:** The likelihood - – of the training data d is therefore:

Taking the log of both sides we get:

* **Maximum Likelihood Estimate (ML):** The ML of maximises the likelihood. Equivalently, it maximises the log-likelihood. We can also drop the scaling factor . This, therefore leaves us with a new equation to me maximised:
* **Maximum Posterior Estimate (MAP):** The MAP is an estimate for that maximises the posterior of Bayes rather than the likelihood.

Consider the following model:

Where:

i.e Gaussian distributed with mean 0 and variance 1

i.e Gaussian distributed with mean 0 and variance

We already know that the likelihood is:

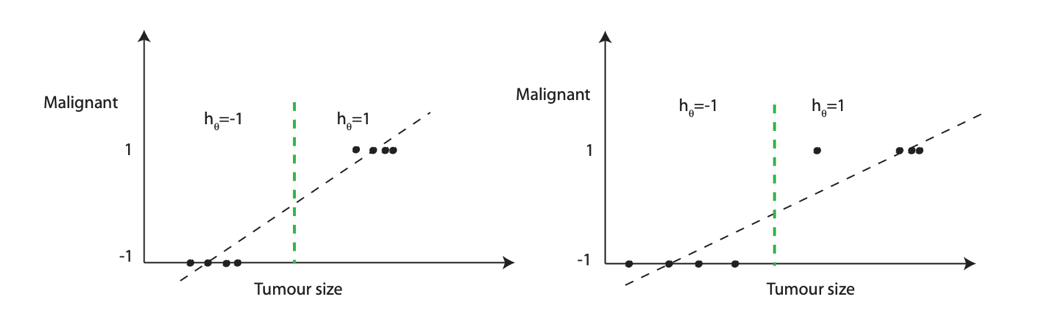
From our model, we have prior as:

The evidence (denominator - ) is a normalising constant, so area under PDF = 1. Combing these rules with Bayes we then have:

The MAP estimation for this given model is the value of that maximises

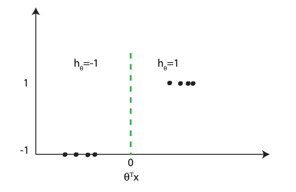
**Logistic Regression**

* **Classification:** Logistic regression serves the purpose of classification (email – spam or not spam?). Y values now only take values {-1, 1} whereas before Y was real-valued. We want to build a classifier that predicts the label of a new object (whether spam or not).

****

Line of best fit no longer works here, can result in misclassification. A more suitable model would be to predict output 1 (malignant) when and output -1 (not malignant) when

* **Plane Fitting:** Instead of trying to choose a line of best fit between data points, logistric regression aims to choose a plane that separates Y=1 data from the Y=0 data.

****

* **Logistic Regression Cost Function:** Similar to how in linear regression we used a cost function such as least squares to find the line of best fit, in logistic regression we use:

Scaling by log(2) here is optional, but it makes the loss 1 when .

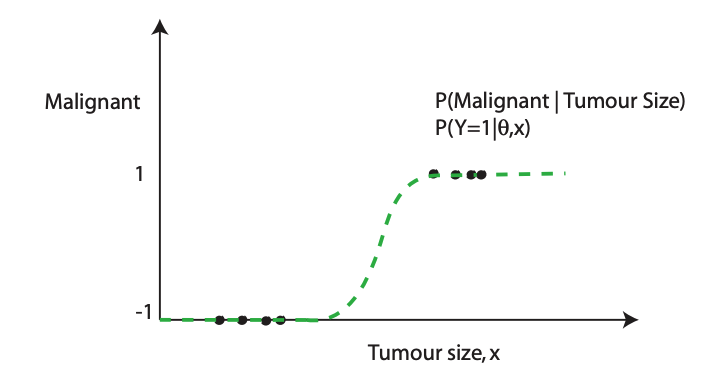
* **Maximum Likelihood Estimate (ML):** Label Y only takes values of -1 and 1. Assume:

The likelihood ) of the training data d is therefore:

Taking logs, this gives us:

The ML can be considered the value of that maximises the above equation. However, in order to simplify this, we can use the log rule and then the ML becomes the value of that **minimises**:

* **Example:** Using the above examples, and the hypothesis , we now have an estimate for our confidence in the prediction, namely:

****

We can see that when is close to 1 then we are confident in our prediction, but when is small then we are less confident in our prediction.