Advanced Regression Notes:

Summarising the important takeaways from the lecture:

1. In generalised regression models, the basic algorithm remains the same as linear regression. We compute the values of coefficients which result in the least possible error (best fit). The only difference is that we now use the features ϕ1(x),ϕ2(x),ϕ3(x)....ϕk(x) instead of the raw attributes.
2. The term 'linear' in linear regression refers to the linearity in the coefficients, i.e., the target variable y is linearly related to the model coefficients. It does not require that y should be linearly related to the raw attributes or features; feature functions could be non-linear as well.

‘Linear’ in Linear Regression:

 The model is called 'linear' because the targety**is linearly related to the** **coefficients**. To fully understand this, it is crucial to note that in regression, the**coefficients** a0,a1,a2,...,ak are your **variables**, i.e., you are trying to find the optimal coefficients that minimise some loss function. On the other hand, the **features** ϕ1(x),ϕ2(x),ϕ3(x)....ϕk(x) are actually **constants** because you are already given the dataset (i.e., the values of x, and hence ϕ(x), are fixed; so, what you are trying to tune are the coefficients).

Thus, saying that 'y is linearly related to the coefficients' implies that **only two operations**can be applied between the coefficients:

1) **Multiplying them by constants** (i.e., the features) such as a1ϕ1(x),a2ϕ2(x) and

2)**Adding the terms**with each other such asa1ϕ1(x)+a2ϕ2(x). What you **cannot** do is multiply them together, raise to one another's power, etc. That is, you cannot have terms such as a0.a1,aa32 etc.

Solving this system means to find the set of parameters a0,a1,...,a5 which satisfies all the equations. This can be done efficiently using **matrices**(and is done that way by many libraries)**.** The alternate way is to use optimisation methods such as gradient descent.

It turns out that the closest approximation of X−1 (for non-square matrices) is (XTX)−1XT. Thus, the (approximate) solution to this system is given by:



