**Generalized linear model (GLM)**

GLM is fitting generalized linear models, specified by giving a symbolic description of the linear predictor and a description of the error distribution.

**Advantages:**

* This model could extend the linear model to problems in which the response is categorical or discrete rather than a continuous numeric variable. Although the response itself may conform to arbitrary distributions, with the help of link function, the predicted values vary linearly
* GLM have the random component in the model. The response variables may have error distribution models other than a normal distribution. If we assume the response has a Gaussian distribution, then the linear models are a special case of GLMs.

**Notations:**

* X: a series of features, which specifies a linear predictor for response.
* Y: the response variable.
* :The parameters of the linear predictor. GLM are typically fit to data by the method of maximum likelihood, using iteratively weighted least squares procedure.

**Model components:**

* A random component.

This random component specifies the conditional distribution of the response variable, y, given the predictors. The conditional distributions are from an exponential family.

P.S. : In R, the glm includes family-generator functions for the five standard exponential families. Each family has its own canonical link, which is used by default if a link is not given explicitly.

* A linear predictor . : The value of the linear predictor. is the estimated value of the linear predictor.
* A link function: The link function, which provides the relationship between the linear predictor and the mean of the distribution function. For example, can take on any value in , whereas the mean of a binary random variable must be in the interval The link function translates from the scale of the mean response to the scale of the linear predictor.is the expected value of the response and is the linear predictor.

**Statistical Analysis:**

Assume the estimated value of the linear predictor as .

* The estimated mean of the response is , because of the link function
* The variance of distribution is

is the positive dispersion parameter related to the variance of the exponential family. For the binomial and passion distributions, the dispersion parameter equal to 1, for Gaussian data, the parameters could be replaced by