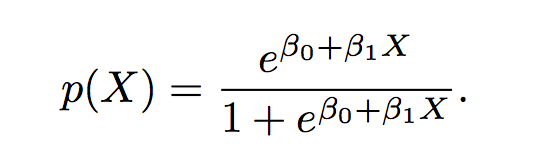
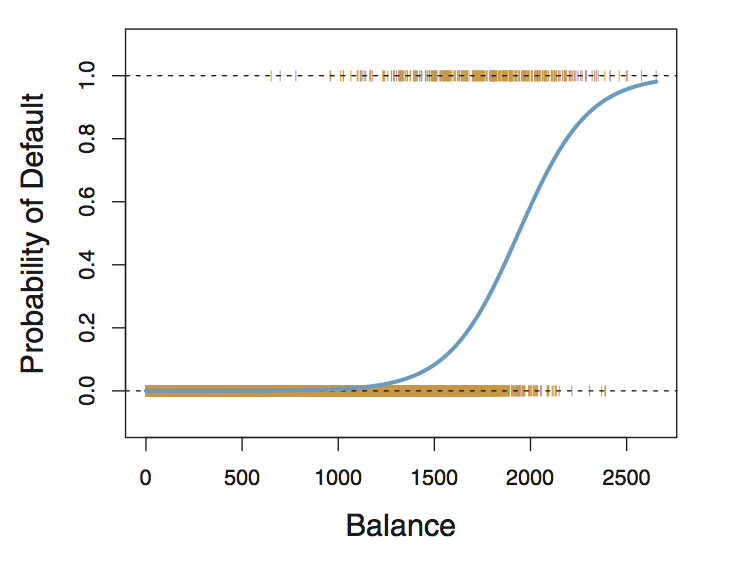
**CHEAT SHEET 2: LOGISTIC REGRESSION**

*Formula:*

****

*Plot*:



*Suitable for*: Predicting the value of a binary categorical variable

*Procedure*: **(1) Explore the relationships** using visualisation and correlation. **(2) Transform variables** into binary variable; **(3) Run the model. (4) Test the model** by fitting it against a training set and seeing how many it can predict.

*Interpreting the model*. **Estimate** is the increase in the logit score when the unit increases by 1; the **standard error** is the 95% confidence interval boundaries (+ or -) of the estimate; the **z value** is the coefficient divided by its standard error; the **p value** gives an indication of whether or not the slope of the line differs significantly from 0; **null deviance** shows how well the response variable is predicted by a variable that includes only the mean; the **residualdeviance** shows how well the model fits with all the features included. For both residual and null deviance, the smaller the better.

*Actions* *for improving models*. **(1) Add or remove variables; (2) Tune the pvalue** for classification of the independent variable.

*Considerations:* **(1) Check for confounding variables.**