Logistic regression: A note on the Matlab code
Linear regression: An example calculation using mean and std. dev. values

Logistic regression is a nonlinear variant of the linear models for regression seen in class. In the linear models, although the basis functions $\Phi(x)$ can be nonlinear, the models are linear on w. So if one plots $\Phi(x_i)$ against y this should be a straight line.

In the case of logistic regression, a nonlinear function can be applied to the linear combination on w, such that:

$$y = f(w,\Phi(x))$$

In Matlab, the *link* function is this function f. In the Matlab example code on logistic regression provided in the tutorial, initially f is “logit”.

Changing the link function to “identity” allows you to simulate the linear models seen in class.

Using a linear basis function as well will produce the simplest form of liner regression.

For example, given a training set with points X and target Y below:

X	Y
1.00	1.00
2.00	2.00
3.00	1.30
4.00	3.75
5.00	2.25

we would like to find values for w_0 and w_1 such that newY = $w_0 + w_1X$ minimizes the sum of the squared errors $(Y – newY)^2$ under Gaussian noise. This can be done using:

$w_1 = \rho\, \sigma_y/\sigma_x$

$w_0 = \mu_y – w_1\, \mu_x$

where ρ is the correlation between X and Y, σ is the standard deviation, and μ the mean. In this example, newY = 0.425 X + 0.785

Now, calculate the training set error (i.e. the mean squared error on the training set above).

Given the test set below, what are the predictions for Y? What is the test set error?

X	Y
1.50	1.40
2.50	2.20