Chapter 1 - Introduction

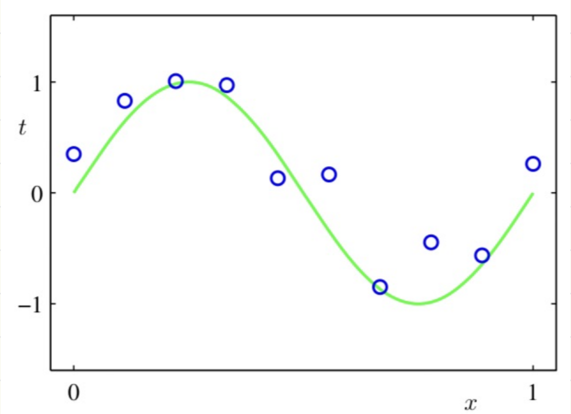
pattern recognition - automatic discovery of regularities in data through the use of computer algorithms
training/learning set - large set of input used to tune the parameters of an adaptive model
test set - new input ran through the trained model
generalisation - ability to categorise correctly new examples that differ from those used for the training
supervised learning - training set contains input with corresponding target
classification - assign each input to one of a finite number of discrete categories
regression - desired output consists of one or more continuous variables
unsupervised learning - training set without targets
clustering - discover groups of similar examples within the data
density estimation - determine the distribution of data
visualisation - project data from high-dimensional to 2/3D space
reinforcement learning - find suitable actions maximising reward
credit assignment - trade-off between exploration and exploitation

1. Example: Polynomial Curve Fitting
t = sin(2πx) + noise
polynomial curve fitting function:

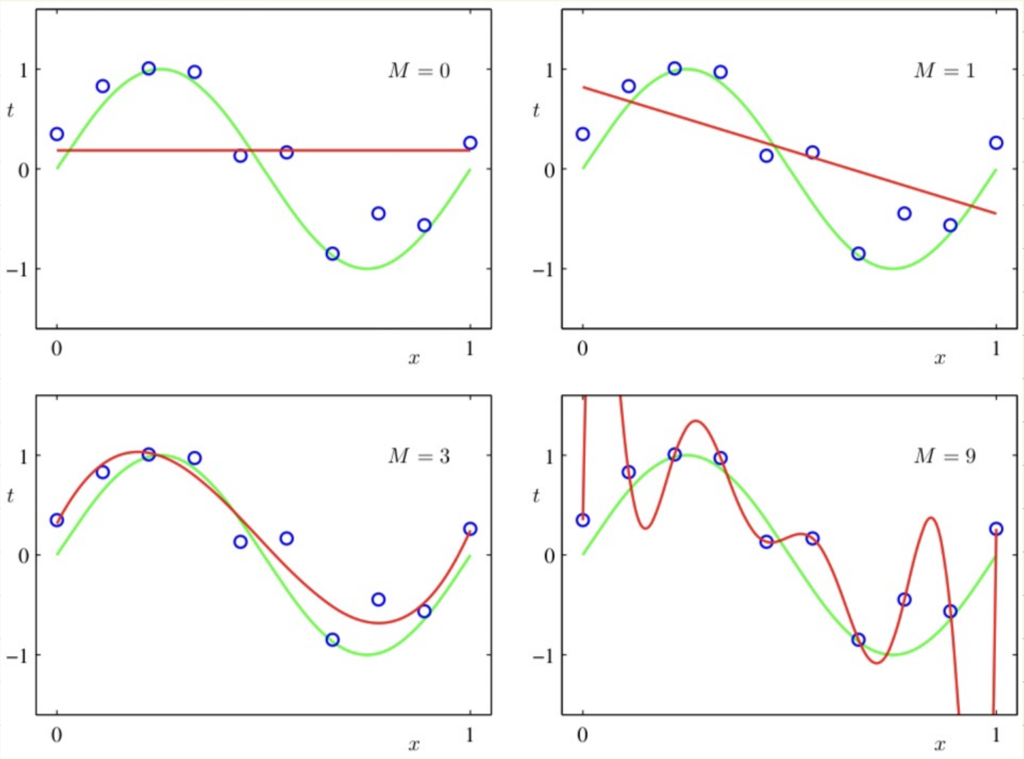
$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$

error function:

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N} \{y(x_n, \mathbf{w}) - t_n\}^2$$



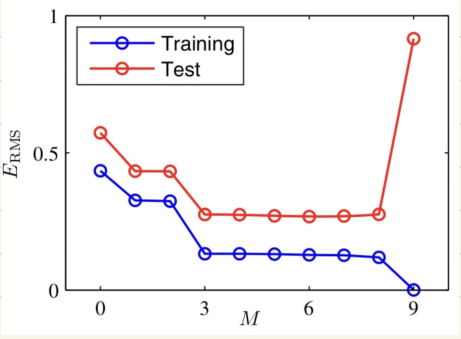
model comparison/selection - problem of choosing the order M of the polynomial
M = 9, E(w*) = 0 - over-fitting (passes precisely through all the points)



root-mean-square error:

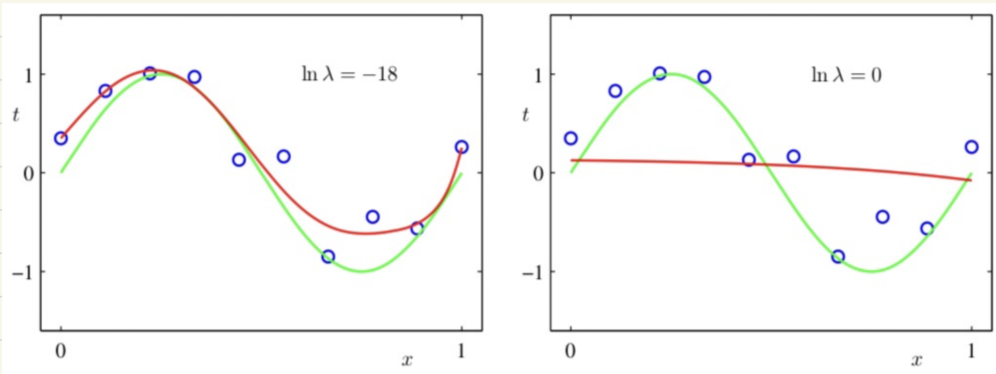
$$E_{\mathrm{RMS}} = \sqrt{2E(\mathbf{w}^\star)/N}$$

	$M=0$	$M=1$	$M=6$	$M=9$
$w_0^\star$	0.19	0.82	0.31	0.35
$w_1^\star$		-1.27	7.99	232.37
$w_2^\star$			-25.43	-5321.83
$w_3^\star$			17.37	48568.31
$w_4^\star$				-231639.30
$w_5^\star$				640042.26
$w_6^\star$				-1061800.52
$w_7^\star$				1042400.18
$w_8^\star$				-557682.99
$w_9^\star$				125201.43



regularisation - controls over-fitting phenomenon

$$\widetilde{E}(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n,\mathbf{w}) - t_n\}^2 + \frac{\lambda}{2}\|\mathbf{w}\|^2$$



	$\ln\lambda = -\infty$	$\ln\lambda = -18$	$\ln\lambda = 0$
$w_0^\star$	0.35	0.35	0.13
$w_1^\star$	232.37	4.74	-0.05
$w_2^\star$	-5321.83	-0.77	-0.06
$w_3^\star$	48568.31	-31.97	-0.05
$w_4^\star$	-231639.30	-3.89	-0.03
$w_5^\star$	640042.26	55.28	-0.02
$w_6^\star$	-1061800.52	41.32	-0.01
$w_7^\star$	1042400.18	-45.95	-0.00
$w_8^\star$	-557682.99	-91.53	0.00
$w_9^\star$	125201.43	72.68	0.01