# C. Bishop Pattern Recognition
## Section 1.0 intro & 1.1 Poly Curve Fitting

the ability to correctly categorise data/examples not seen in the training data is known as generalization

for most practical applications preprocessing is applied to the data w/ the goal is making the problem easier to solve - i.e. less computationally expensive

test data must also be preprocessed

Bounding Boxes? (Viola & Jones, OK)

Input & target = Supervised learning

input to finite categories = classification

input to continuous = Regression

input w/out targets to learn from = unsupervised, i.e clustering

reinforcement learning (Sutton & Barto, 98)
- find actions to maximize reward and the end
- Discover targets through trial and error

reinforcement learnig Backgammon (Tesauro, 1994)

# 1.1. Polynomial curve fitting

for experiement purposes it is often better to create synthetic data, this way you known the exaction function you wish to find

e.g func $\sin(2\pi x)$ + noise from gaussian distribution

Polynomial function form:

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

$\hookrightarrow$ vector of polynomial coeff, $w_M$

values of coeffs found by minimize a loss function

common = sum of squared errors

$$E(w) = \frac{1}{2} \sum \{ y(x_n, w) - t_n \}^2$$

w is solved by min $E(w)$

but num of M still needs to be determined

this is known as model comparison or model selection

low M will often perform bad & High will appear to fit well but will overfit to the train data & generalize poorly to unseen data

regularization penalize ~~of~~ overfitting & over complicated models

Discourages coeffs reaching large values

$$\tilde{E}(w) = \frac{1}{2}\sum\{y(x_n, w) - t_n\}^2 + \frac{\lambda}{2}||w||^2$$

$||x||$ = ~~eudencidean~~ euclidean Norm

$\sqrt{\text{sum of squared } x}$ or $sqrt(w_0^2 + w_1^2 + w_n^2)$

often $w_0$ coeff is ignored by regularizer because its choice = depends on origin for target (Hastie et al 2001)

or have its own unique regularization factor

Techniques like this are also known as Shrinkage

R: Quadratic regularizer = Hoerel & Kennard, 1970

with neural networks this is known as weight decay

Regularization controls over-fitting and as a result the degree of overfitting