$(x^{(1)}, t_1)\ (x^{(2)}, t_2)\ \cdots\ (x^{(n)}, t_n)$

$x = (x_1, x_2, \cdots, x_d)$

$y(x, \omega)$

$\hookrightarrow y = \omega_1 \phi_1(x) + \omega_2 \phi_2(x) + \cdots \omega_m \phi_m(x) + b$

$y = \omega_1 x_1 + \omega_2 x_2 \cdots \omega_m x_m$

y is linear in ϕ (matured features)

but not in x

SVM

$\underbrace{\begin{matrix} x_1 & x_2 & \cdots & x_d \\ \phi_1 & \phi_2 & \cdots & \phi_m \end{matrix}}_{\text{you may not know}}$

$k(x^i, x^j) = e^{-\frac{\|x^{(i)} - x^{(j)}\|^2}{2\sigma^2}}$

Using SVMs with a kernel $\Rightarrow$ each vector is a support vector $\Rightarrow$ overfit

In SVMs, we fix our features

DL $\Rightarrow$ we get a ton of features

$\Rightarrow$ we dunno how many are appropriate

Problem is, architecture is mostly trial & error