

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

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

$$y(x, w)$$

$$\hookrightarrow y = w_1 \phi_1(x) + w_2 \phi_2(x) + \dots + w_m \phi_m(x) + b$$

$$y = w_1 x_1 + w_2 x_2 + \dots + w_m x_m$$

y is linear in ϕ (mapped features)

but not in x

SVM

$$x_1, x_2, \dots, x_d$$

$$\underbrace{\phi_1, \phi_2, \dots, \phi_m}_{\text{you may not know}}$$

$$k(x^i, x^j)$$

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

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