# Non-Linear Models

① Stay with a linear model, but transform our data.
↳ kernel methods.

② Make the model non-linear
↳ neural networks.

---

Basis function expansion

| $x$ | $y$ |
|---|---|
| 3 | 1 |
| 7 | 3 |
| 4 | -2 |
| 9 | 1.4 |
| 11 | 2.8 |

| ~~$x$~~ $x^2$ | $y$ |
|---|---|
| 9 | 1 |
| 49 | 3 |
| 16 | -2 |
| 81 | 1.4 |
| 121 | 2.8 |

→ ~~$w^T x = y$~~

$w^T x^2 = y$

---

Regularization

Simple → Complex

Simple ↓ larger training

Complex ↓ low training error ≠ low test error

Occam's Razor

How do we force regularization?

Smaller values for the weights ⇒ less complex models.

| | $x$ | $x^2$ | $x^3$ | |
|---|---|---|---|---|
| $w_0$ | $w_1$ | $w_2$ | $w_3$ | |
| 2 | 7 | 9 | 3 | → more complex |
| 1 | 3.5 | 4.5 | 1.5 | → |

## Linear regression

$$J(w) = \frac{1}{2}(y - Xw)^T (y - Xw)$$

error on the training data.

$$\boxed{\bar{J}(w) = J(w) + \frac{1}{2}\lambda \|w\|^2}$$

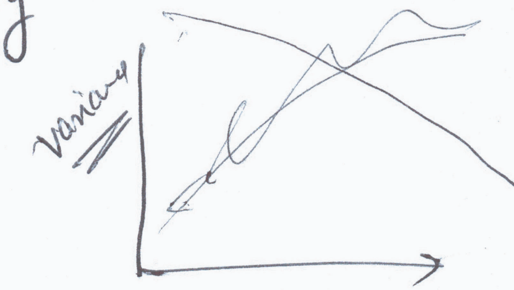
l2 norm

$\|w\|^2 = (w_1^2 + w^2 + \dots)$

Regularization Penalty

a scalar value

$0 \leq \lambda \dots$




## Ridge Regression

$$\bar{J}(w) = J(w) + \lambda |w|$$

LASSO

---

$$\frac{d}{dw}\bar{J}(w) = \frac{d}{dw}\left[\frac{1}{2}(y - Xw)^T (y - Xw)\right] + \frac{\lambda}{2}\frac{d}{dw}\|w\|^2$$

$\|w\|^2 = w^T w$

$$= X^T X w - X^T y + \lambda w$$

$$\frac{d}{dw}\bar{J}(w) = 0$$

$$X^T X w - X^T y + \lambda w = 0$$

$$\boxed{w = (X^T X + \lambda I)^{-1} X^T y}$$

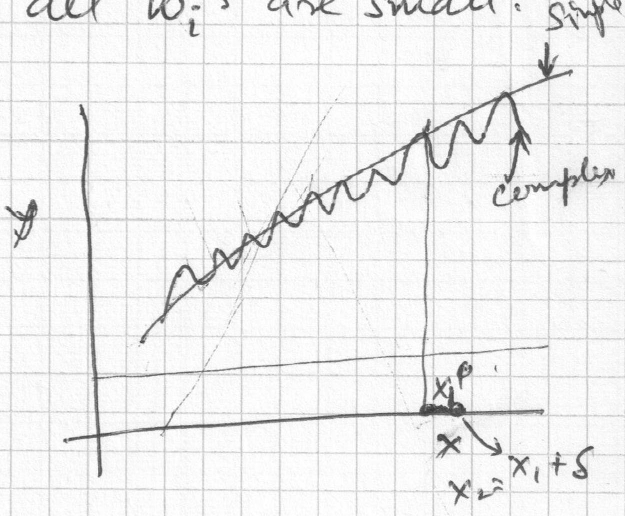
Identity matrix

$D \times D$

$$\begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$$

Why does a small $\|W\|_2^2$ ensure a simpler model?

Essentially means that all $w_i$'s are small.

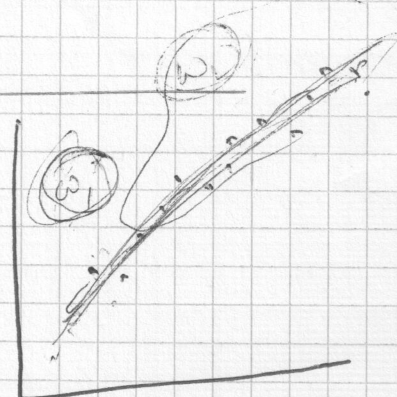



$y = w_0 + w_1 x$

$y_1 = w_0 + w_1 x_1$

$y_2 = w_0 + w_1 x_2$
$= w_0 + w_1 (x_1 + \delta)$

$y_2 - y_1 = w_1 \delta$




$y = w_0 + w_1 x + w_2 x^2$

$y_1 = w_0 + w_1 x_1 + w_2 x_1^2$

$y_2 = w_0 + w_1 (x_1 + \delta) + w_2 (x_1 + \delta)^2$

$$\boxed{y_2 - y_1 = w_1 \delta + 2 w_2 x_1 \delta + w_2 \delta^2}$$

Ridge

$\|W\|_2^2$

$(w_1^2 + w_2^2 + w_3^2)$

$\downarrow$

$(0 \quad 0 \quad \dots)$

LASSO

$|W|$

$(|w_1| + |w_2| + |w_3|)$

$0 \quad 0 \quad \dots$