

Soit  $L$  l'erreur d'apprentissage totale.

$$L = \sum_{(x_t, y_t) \in D_{train}} (f(x_t) - y_t)^2$$

Soit  $R$  le critère de régularization.

$$R = \lambda \|w\|^2 = \lambda \sum_{i=1}^d w_i^2$$

Soit  $C$  le critère total qui est minimisé par notre algorithme d'apprentissage:

$$C = \lambda \|w\|^2 + \sum_{(x_t, y_t) \in D_{train}} (f(x_t) - y_t)^2$$

Notez que

- $\tilde{X}\theta$  est un vecteur (colonne) qui contient la prédiction  $f(x_t)$  pour chaque rangée de  $X$ ;
- $w = I_d \theta$ .

Mettre  $C$  en terme de  $\theta$ ,  $\tilde{X}$  (on a besoin de différencier par rapport à  $\theta$ ). La cible est un polynôme de degré 2 en  $\theta$ .

$$C = \lambda \|w\|^2 + \sum_{(x_t, y_t) \in D_{train}} (f(x_t) - y_t)^2 \quad (1)$$

$$= \lambda \|w\|^2 + \sum_{(x_t, y_t) \in D_{train}} (b + w'x_t - y_t)^2 \quad (2)$$

$$= \lambda \|w\|^2 + \|\tilde{X}\theta - Y\|^2 \quad (3)$$

$$= \lambda w'w + (\tilde{X}\theta - Y)'(\tilde{X}\theta - Y) \quad (4)$$

$$= \lambda(\theta I_d)'(I_d \theta) + (\tilde{X}\theta - Y)'(\tilde{X}\theta - Y) \quad (5)$$

$$= \lambda(\theta I_d)'(I_d \theta) + \theta' \tilde{X}' \tilde{X} \theta - 2Y' \tilde{X} \theta + Y' Y \quad (6)$$

$$= \theta'(\lambda I_d + \tilde{X}' \tilde{X})\theta - 2Y' \tilde{X} \theta + Y' Y \quad (7)$$

Finalement, on dérive par rapport à  $\theta$ . (Notez que  $\frac{\delta}{\delta x} Wx = W'$ .)

$$\frac{\delta C}{\delta \theta} = 2(\lambda I_d + \tilde{X}' \tilde{X})\theta - 2\tilde{X}' Y \quad (8)$$

La solution  $\frac{\delta C}{\delta \theta} = 0$  est un minimum parce que

1. la matrice  $\tilde{X}' \tilde{X}$  est définie positive quand  $T \geq n$  et les  $x$  sont tous linéairement indépendants,
2. la matrice  $\lambda I_d$  est semi-définie positive,

3. la somme de ces deux matrices est nécessairement définie positive,  
et alors  $C(\theta)$  est un polynôme de degré 2 avec une matrice quadratique qui est définie positive.  
Au point minimum,  $\frac{\delta C}{\delta \theta} = 0$ . Cela implique que

$$(\tilde{X}'\tilde{X} + \lambda I_d)\theta = \tilde{X}'Y \tag{9}$$