[redacted]

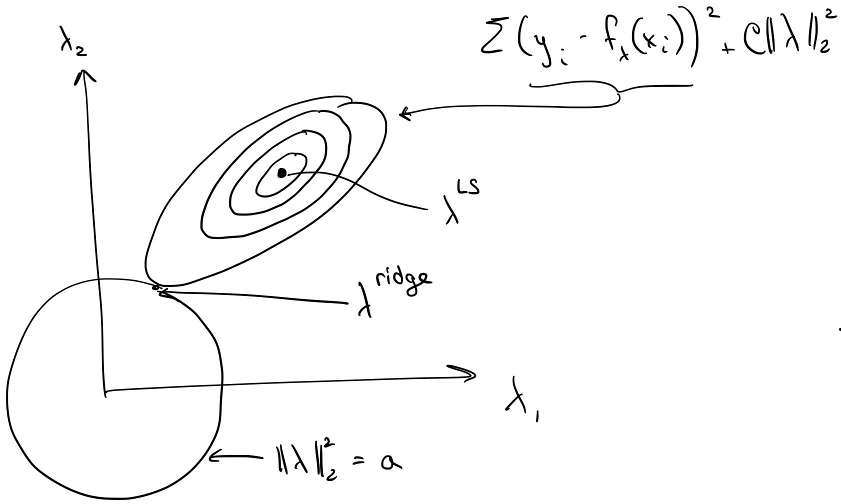
Friends of Ridge Regression

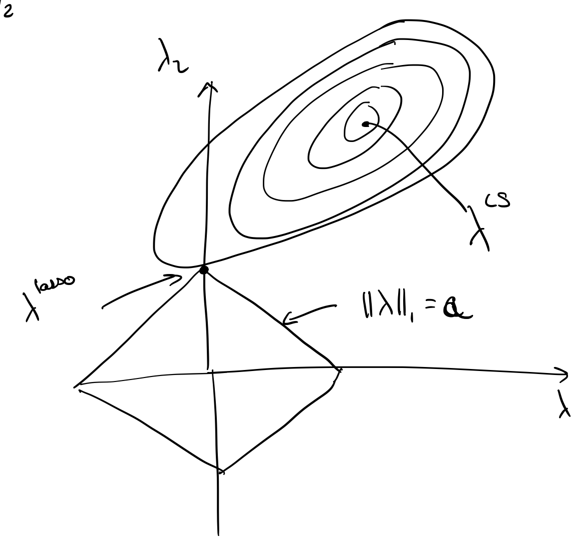
[redacted]

Fact: Lasso does not have a closed form solution.

$$\sum_i (y_i - f_\lambda(x_i))^2 + C\|\lambda\|_1$$

Leads to sparser solutions though





Kernel Least Squares

Regular ridge:

$$F(\bar{\lambda}) = \|\bar{Y} - \bar{\bar{X}}\bar{\lambda}\|_2^2 + C\|\lambda\|_2^2$$

$\bar{\lambda} \rightsquigarrow \bar{\bar{X}}^T \bar{d}$ ← If I can get d, I can get λ.

$$F(\bar{d}) = \|Y - \underbrace{\underset{n\times p}{X}\,\underset{p\times n}{X^T}}_{n\times n} d\|_2^2 + C\|X^T d\|_2^2$$

$$F(\bar{d}) = \|Y - Kd\|_2^2 + C\underbrace{\langle X^T d, X^T d\rangle_2}_{(X^T d)^T (X^T d)}$$

$$d^T \underbrace{X^{\cancel{TT}} X^T}_{K} d$$

$$\langle d, Kd\rangle$$

$$F(d) = \|Y - \underset{n\times n}{K}\,\underset{n\times 1}{d}\|^2 + C\langle d, Kd\rangle$$

$$\nabla F(d) = -2K(Y - Kd) + C2Kd = 0$$

$$-Y + Kd^* + Cd^* = 0$$

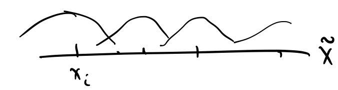
$$d^* = (K + CI)^{-1} Y$$

$X^T = [x_i \,|\,|\,|\,|]$

and remember $\lambda^* = X^T d^* = \sum_i x_i^T d_i^*$

Make prediction at new test point $\tilde{x}$

$$f(\tilde{x}) = \tilde{x} \lambda^* = \tilde{x} \sum_i x_i^T d_i^* = \sum_i \tilde{x} x_i^T d_i^* = \underbrace{\sum_i k(\tilde{x}, x_i)\, d_i^*}_{\text{representer theorem told us this would happen !!}} = K_{\tilde{x}}^T \cdot d^* = K_{\tilde{x}}^T (K + cI)^{-1} y$$

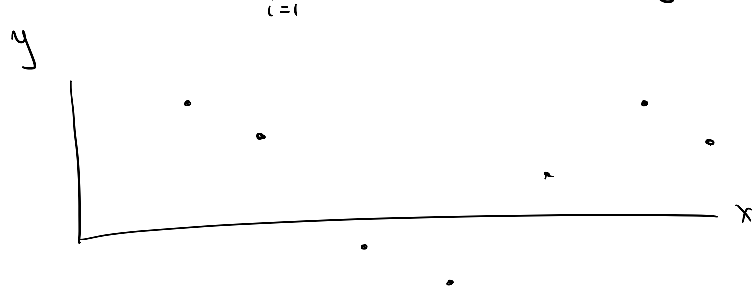


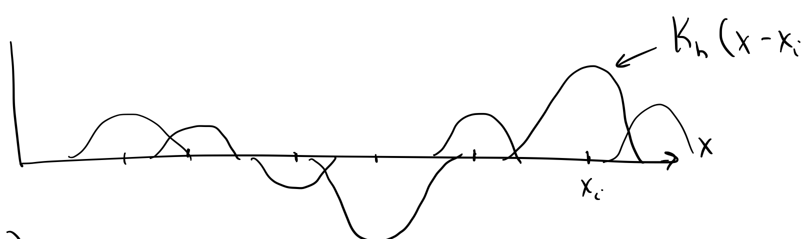
Kernel Regression

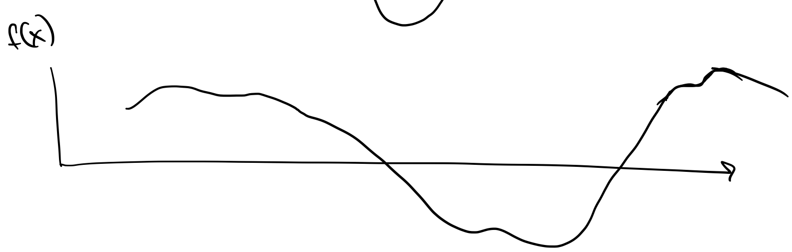
- there is a separate technique called "kernel regression" that is totally different! It is a locally weighted average:

$$f(\tilde{x}) = \frac{\sum_{i=1}^{n} K_h(\tilde{x} - x_i)\, y_i}{\sum_{i=1}^{n} K_h(\tilde{x} - x_i)}$$

← Nadaraya-Watson estimator







- Kernel ridge is certainly more sophisticated - actually tries to minimize least square loss