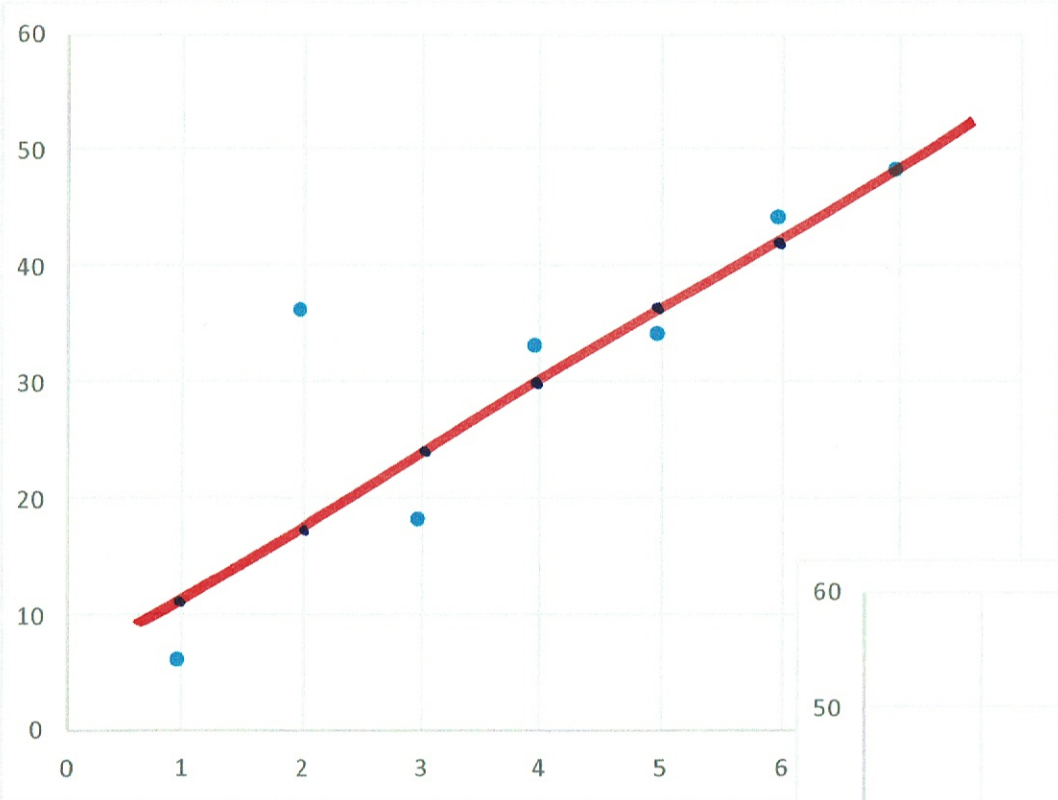
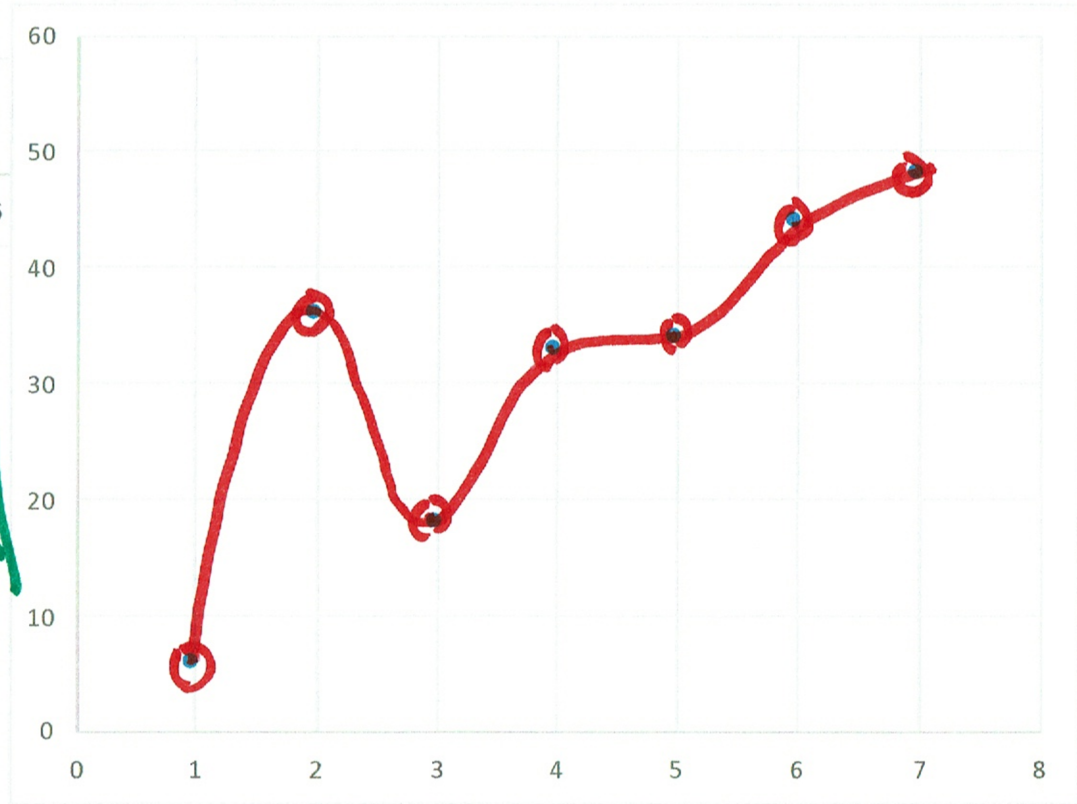
Regularization

Data = Inf + Noise

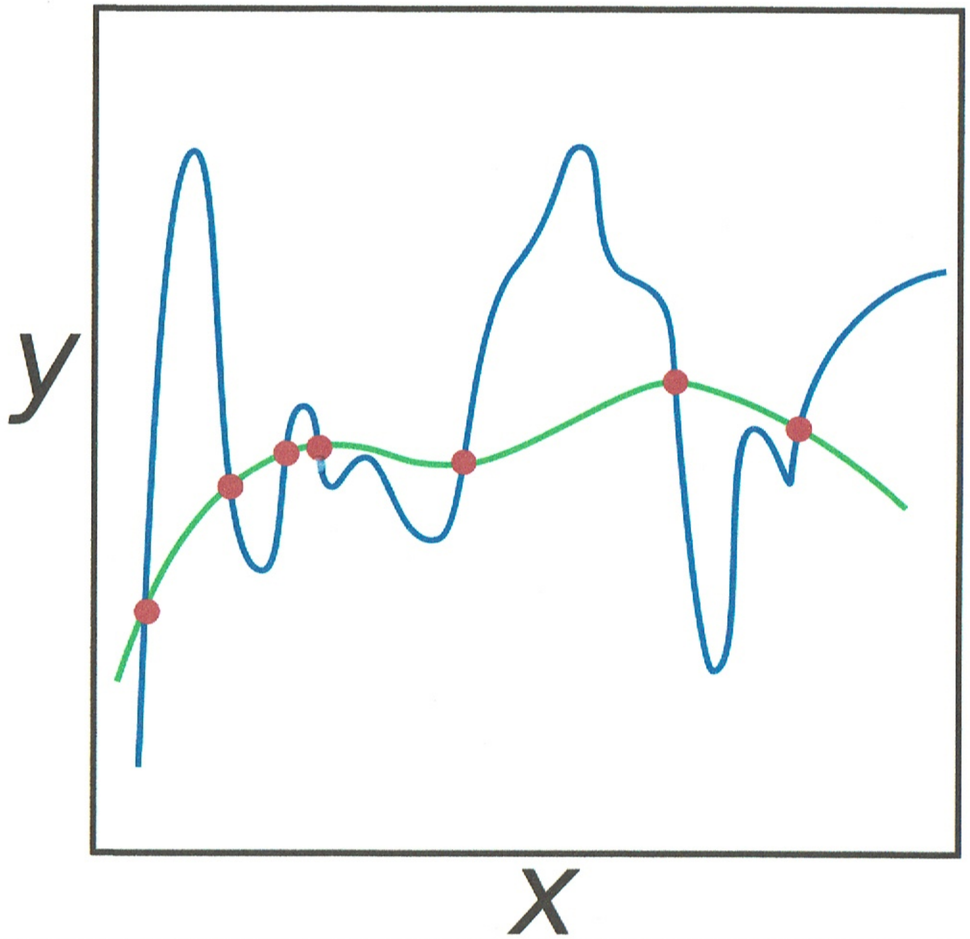


min Loss (error) + λ Penalty for complexity

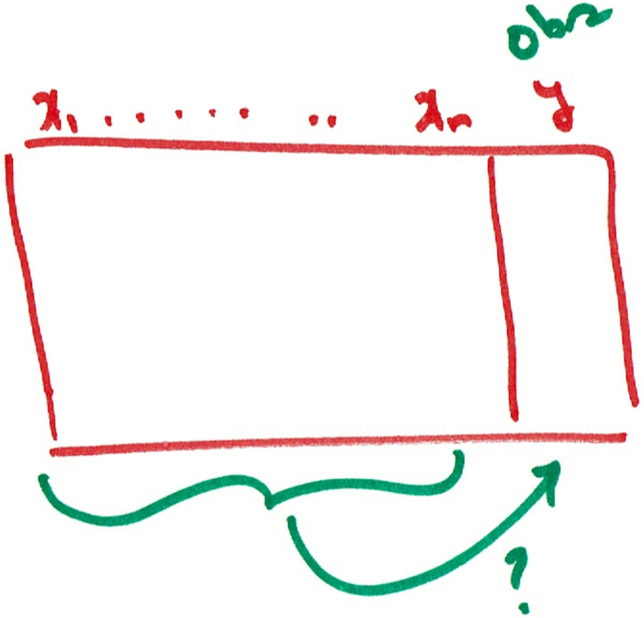


Simpler is better

- "the simplest explanation is most likely the right one" (Occam's razor, Law of Parsimony)
- In Statistics and ML, Regularization prevents overfitting by setting up a preference for simple models



Source: Wikipedia



$y = f(x_1, \ldots \ldots x_n)$

$\hat{y} = w_1 x_1 + w_2 x_2 + \ldots + w_n x_n$

Find w such that

~~$\hat{y}$ & y are close~~

to min $\boxed{(\hat{y} - y)^2 + \lambda \|w\|_p}$

$$\boxed{y = 10x_1 + 20x_2 + 25x_3 + 45x_n}$$

$$\boxed{y = \cancel{25x_2}\ 12x_2}$$

Regularization in Linear Regression

- Accomplished by adding a penalty for complexity!
- The common penalties are of the p-norm type
  - The 1-norm is used in LASSO Regression (induces sparsity)
  - The 2-norm is used in Ridge Regression

$\boxed{\|w\|_p}$ p-norm

$$\|w\|_p = \left(|w_1|^p + |w_2|^p + \dots + |w_n|^p\right)^{1/p}$$

2-norm

$$\|w\|_2 = \sqrt{w_1^2 + w_2^2 + \dots w_n^2}$$

→ **Ridge Reg**, L_2 Reg, Tikhonov

1-norm

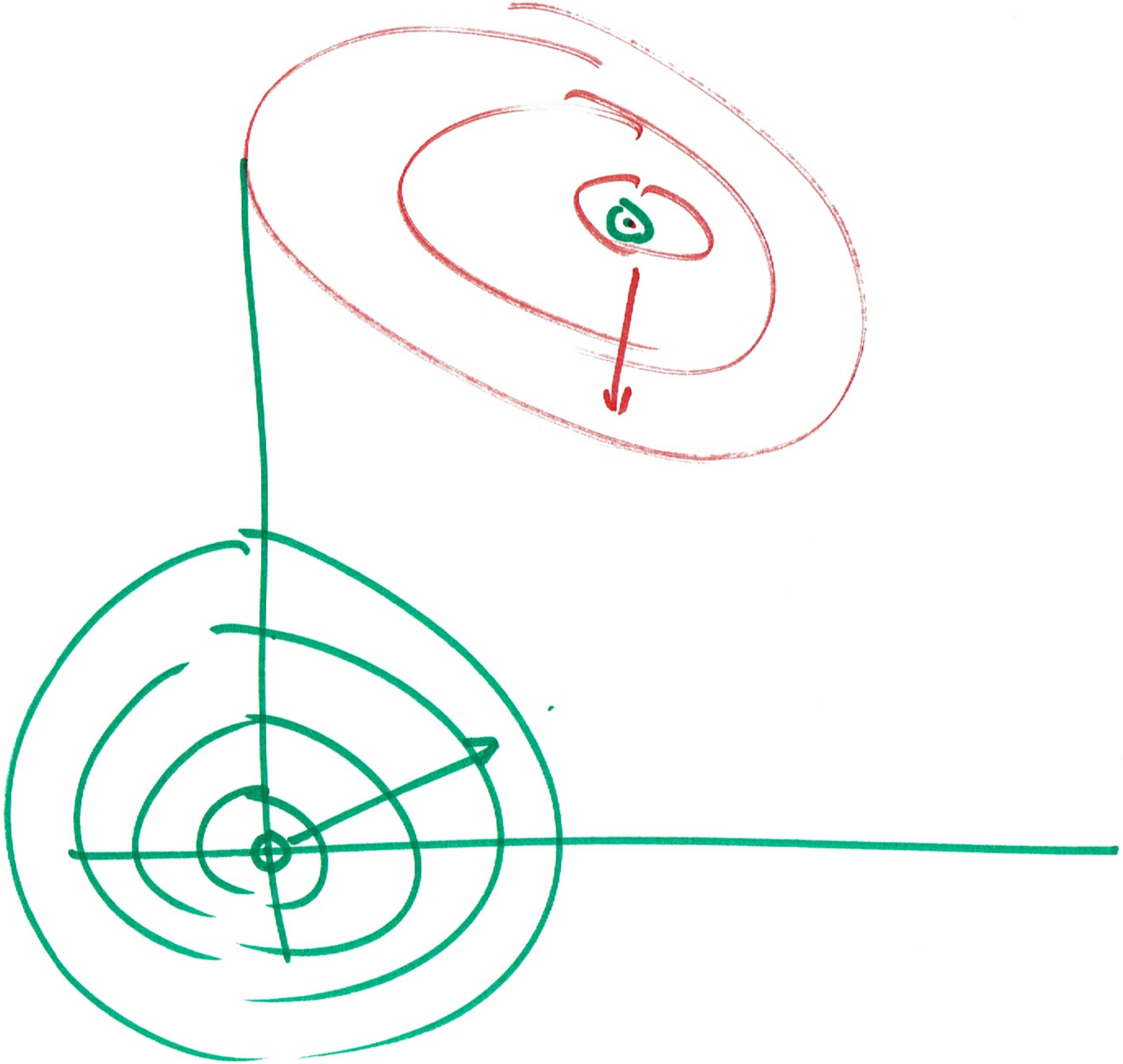
$$\|w\|_1 = |w_1| + |w_2| + \dots + |w_n|$$

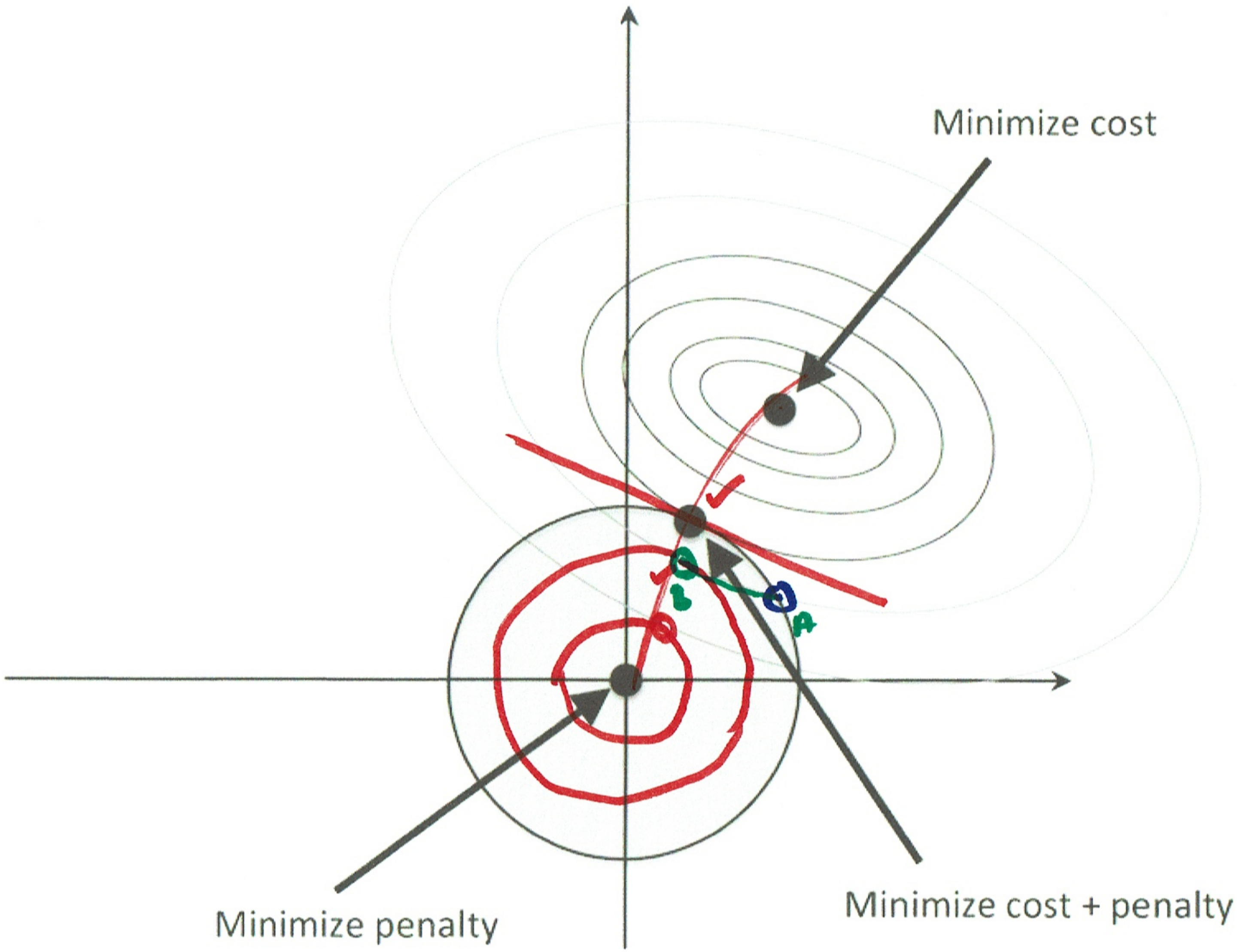
→ **LASSO**, L_1 Reg

L_2 Reg : $\min \underline{\underline{\sum (\hat{y}_i - y_i)^2}} + \underline{\underline{\lambda \sqrt{\sum w_i^2}}}$

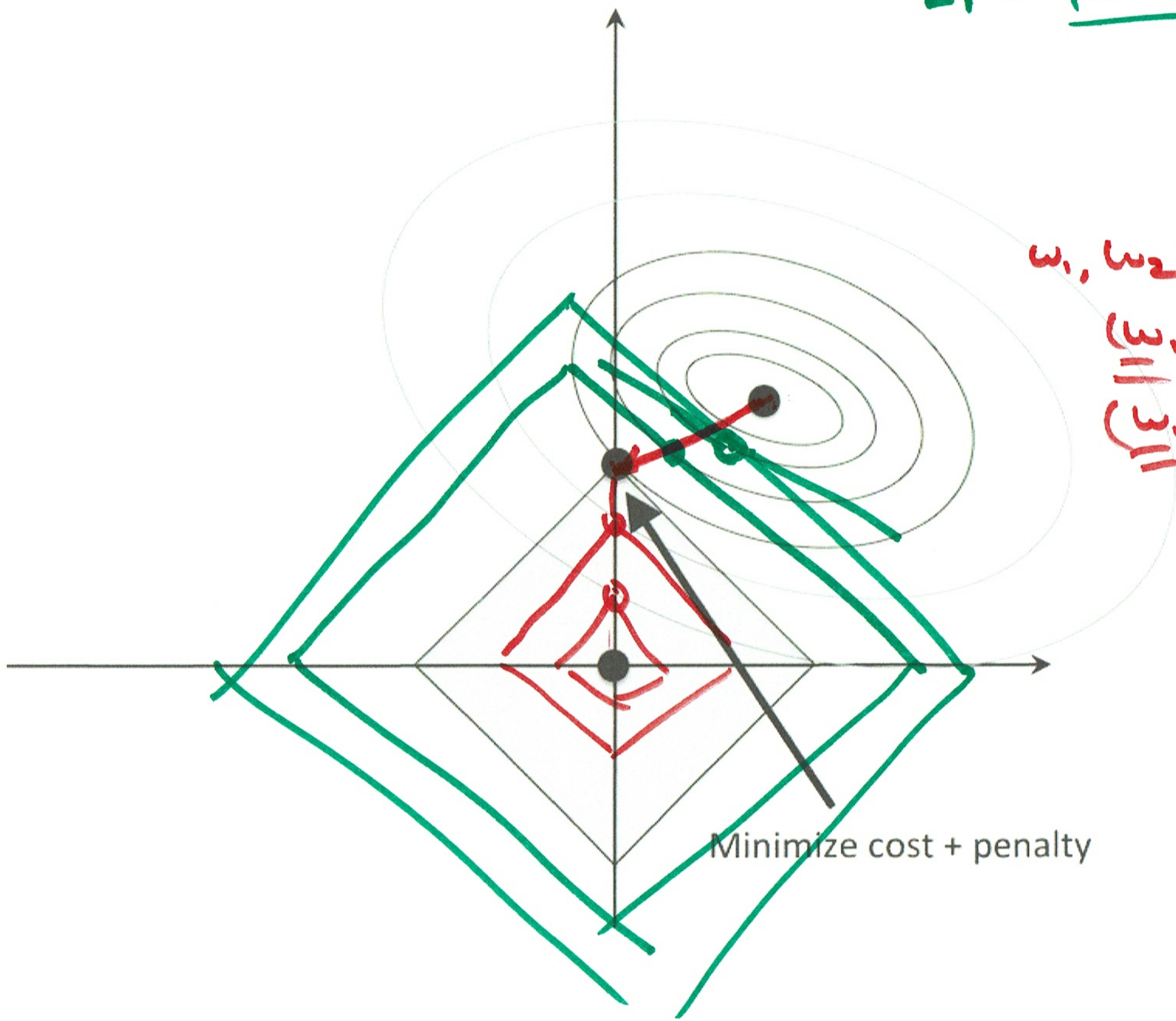
$\hat{y} = w_1 x_1 + w_2 x_2$

$\sqrt{\sum w_i^2} = k$




Minimize cost
Minimize penalty
Minimize cost + penalty

$L_1 \Rightarrow \boxed{\sum |w_i| = 1}$



$w_1, w_2 = 10, 20$

$(w_1, w_2) = 0, 12$

$(w_1, w_2) = 0, 10$

$\vdots$