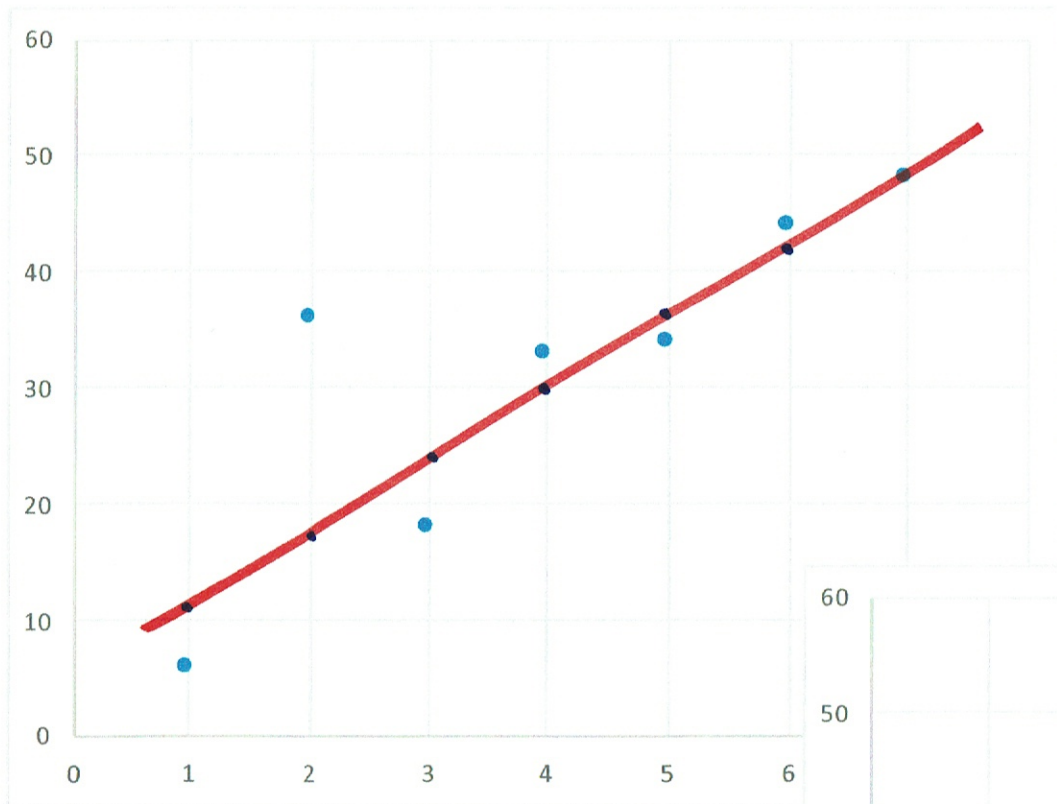
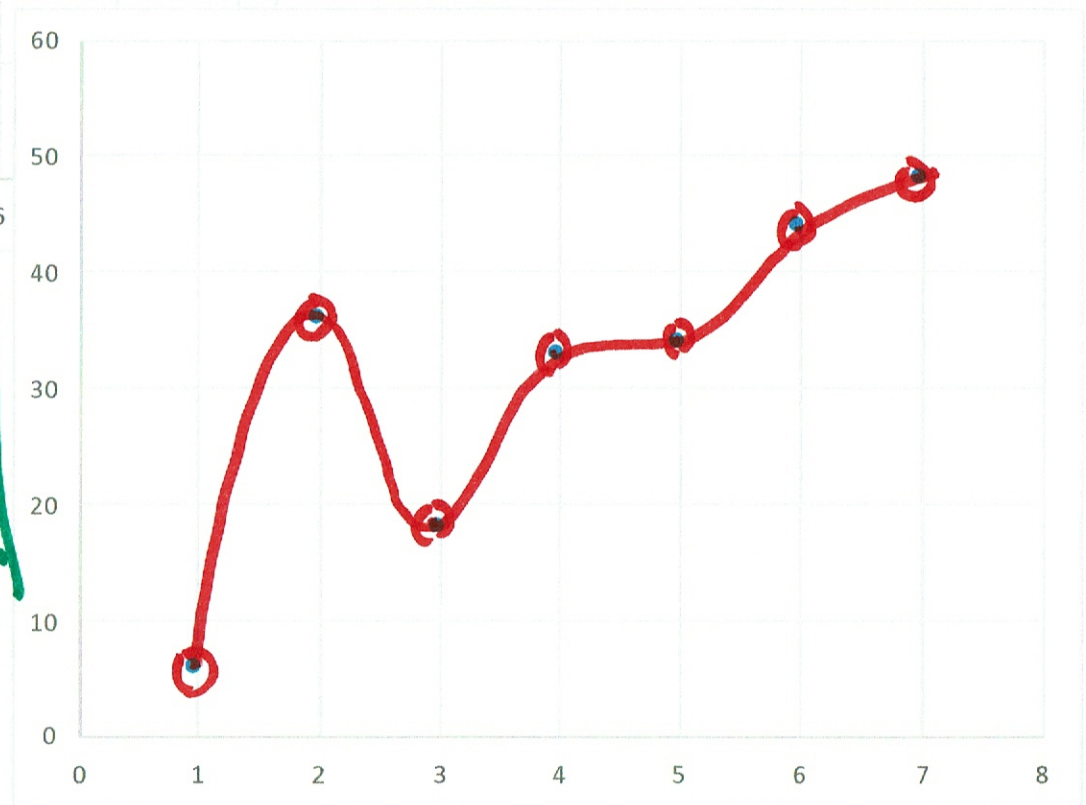


# Regularization



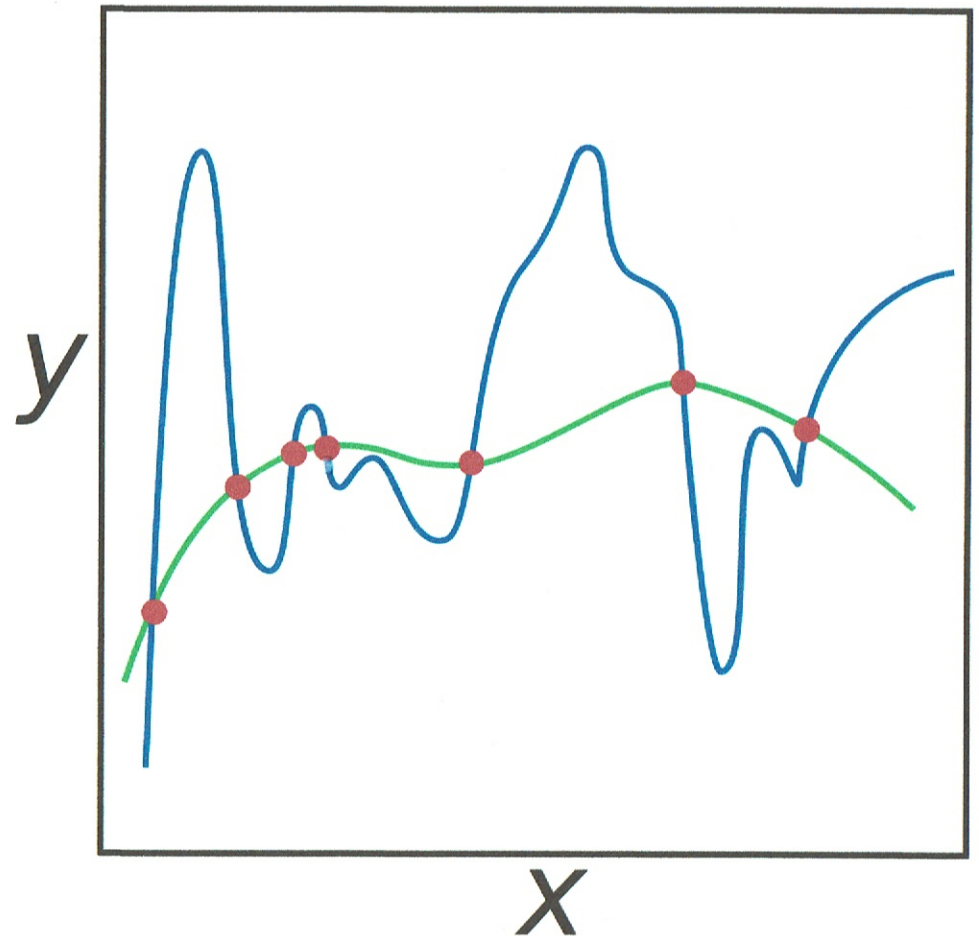
$$\text{Data} = \text{Inf} + \text{Noise}$$

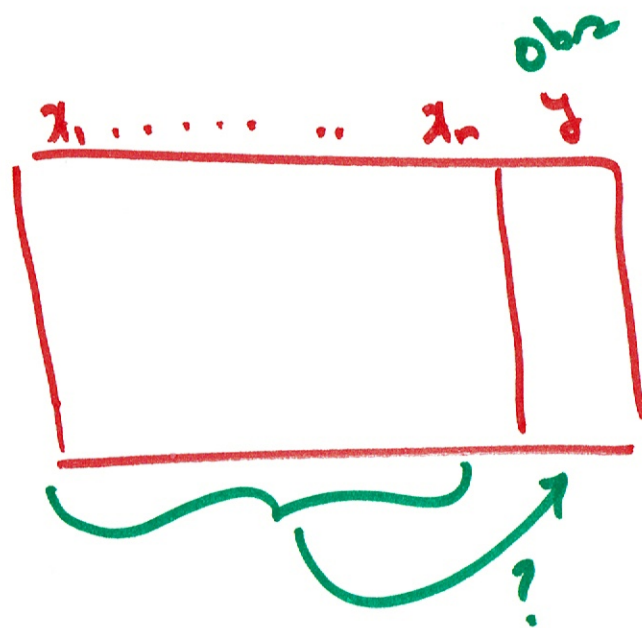
Min  $\left[ \text{Loss (error)} + \lambda \text{Penalty for complexity} \right]$



# 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





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

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

Find  $w$  such that

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

to  $\min \left[ (\hat{y} - y)^2 + \lambda \|w\|_p \right]$

$$y = 10x_1 + 20x_2 + 25x_3 + 45x_4$$

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

# 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



$$\|w\|_p$$

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

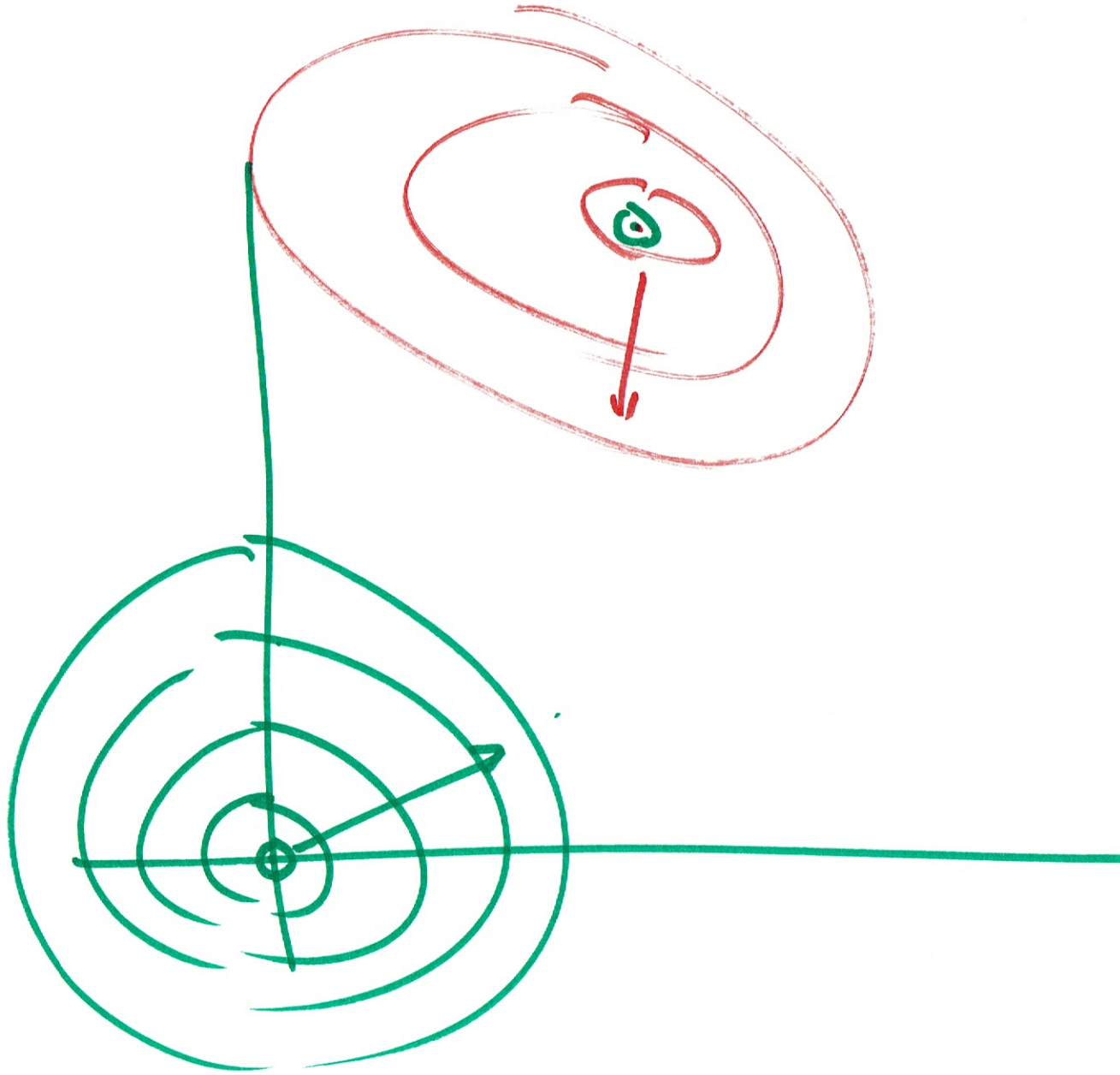
$$\|w\|_1 = |w_1| + |w_2| + \dots + |w_n|$$

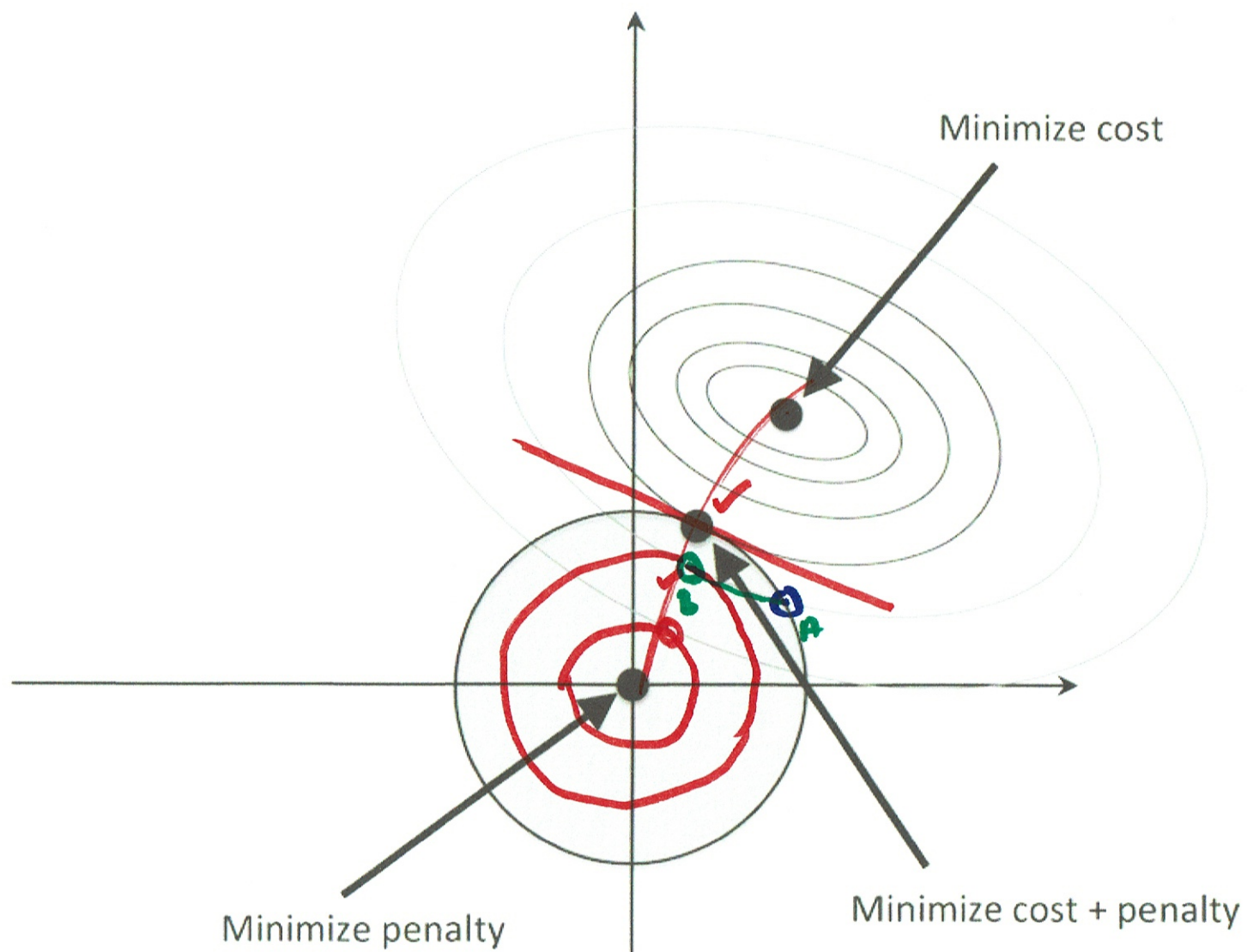
→ LASSO,  $L_1$  Reg

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

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

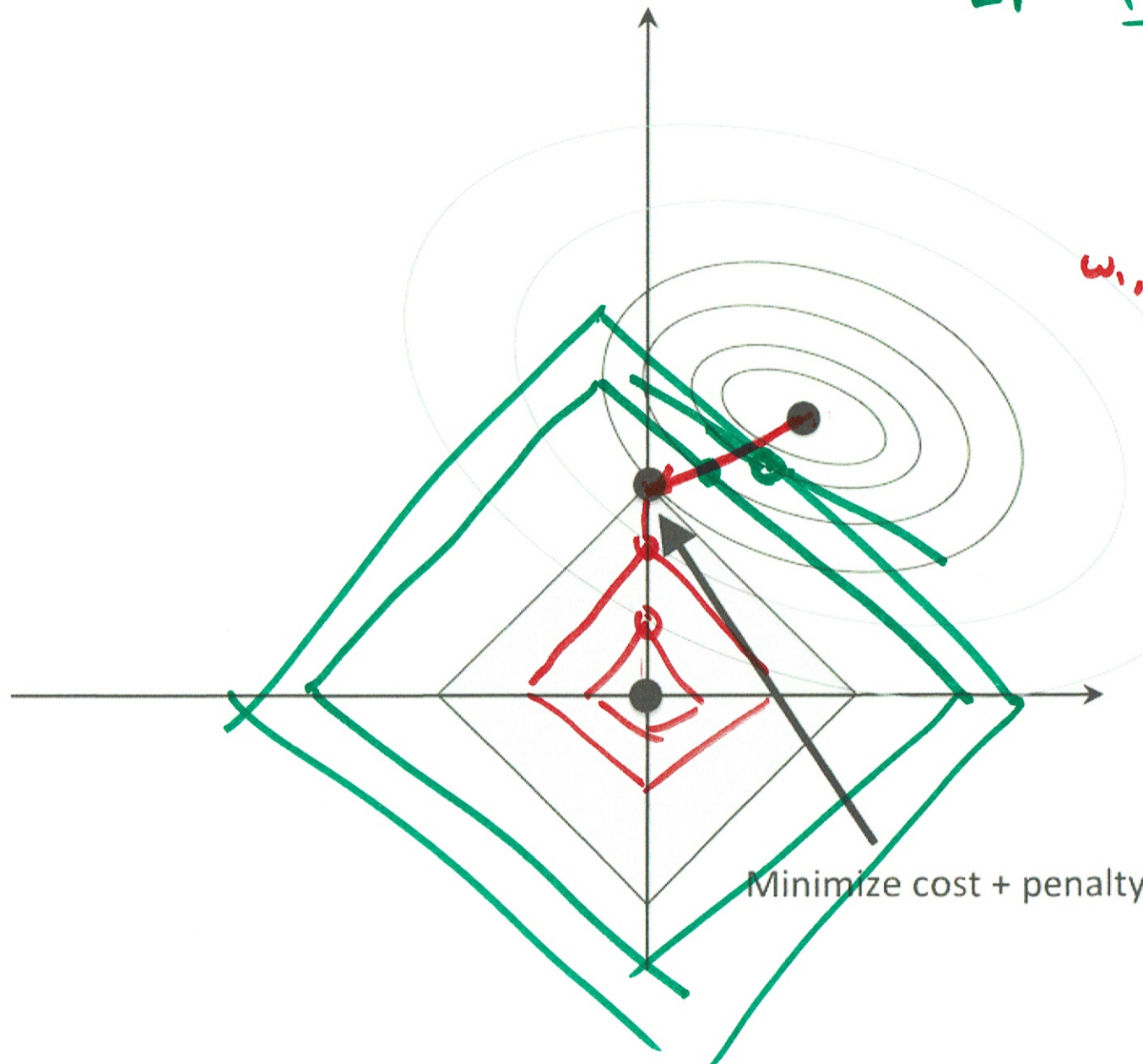
$$\sqrt{\sum w_i^2} = 1/k$$







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



$$w_1, w_2 = 10, 20$$

$$(\underline{w}_1, \underline{w}_2) = 0, 12$$

$$(\underline{w}_1, \underline{w}_2) = 0, 10$$

...

Minimize cost + penalty