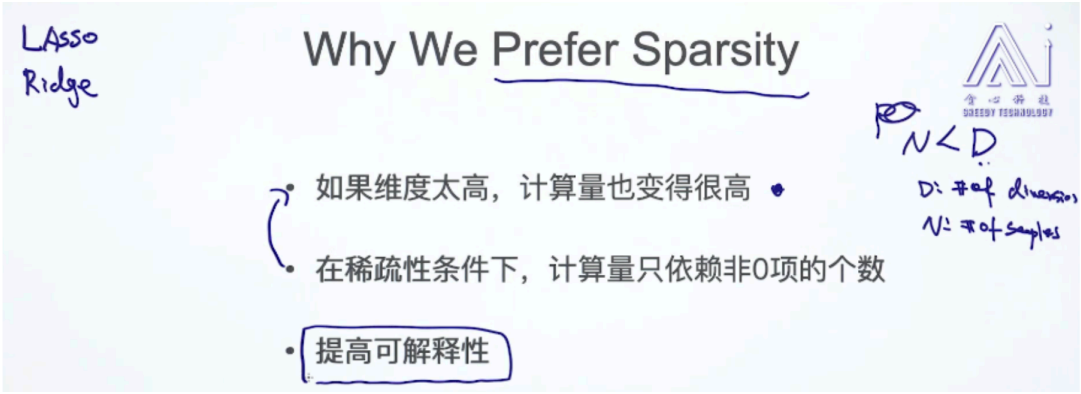
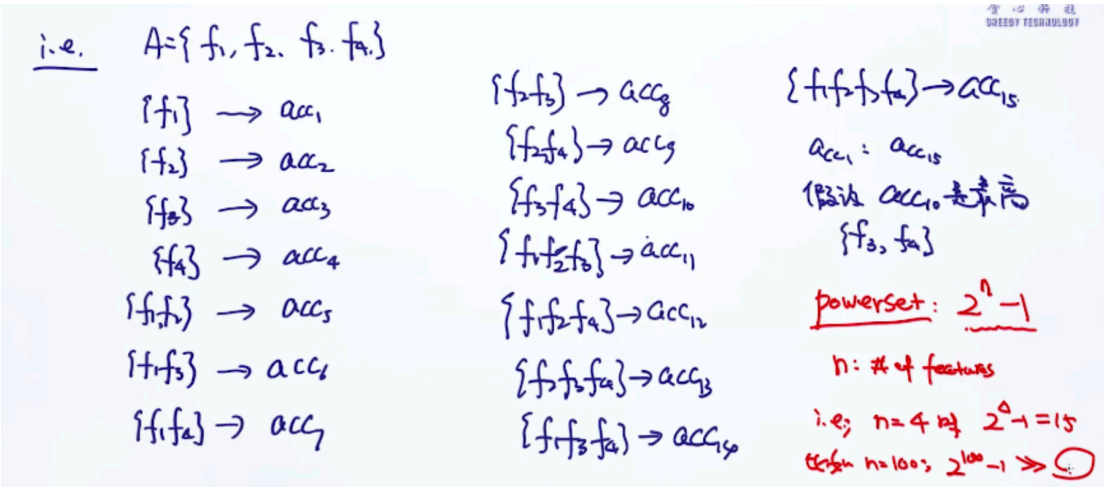
1. Lasso



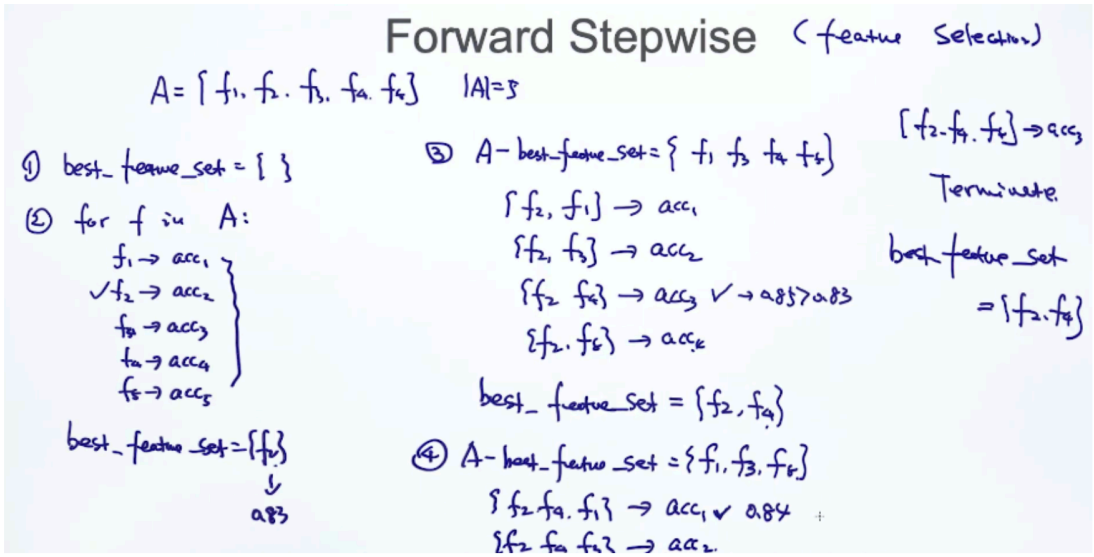
1.1 特征选择

Option1: Exhaustive Search : “all subsets”

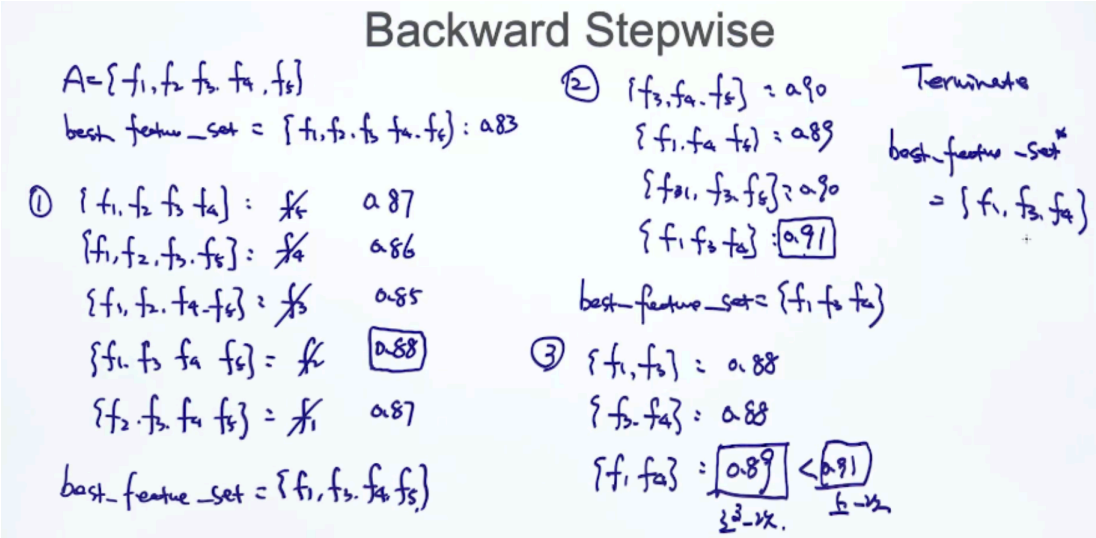


Option2: Greedy Approaches

Forward Stepwise & Backward Stepwise

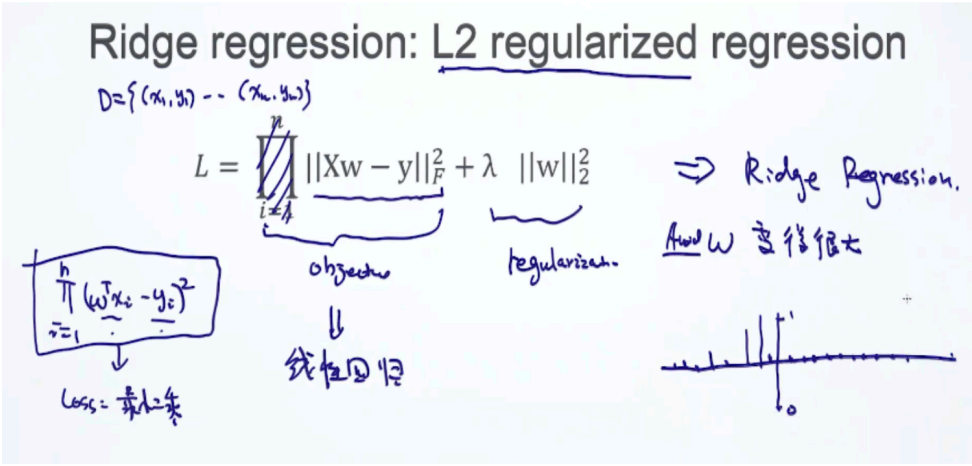


只能是局部最优解，因为我们只能看到目前最好的情况。

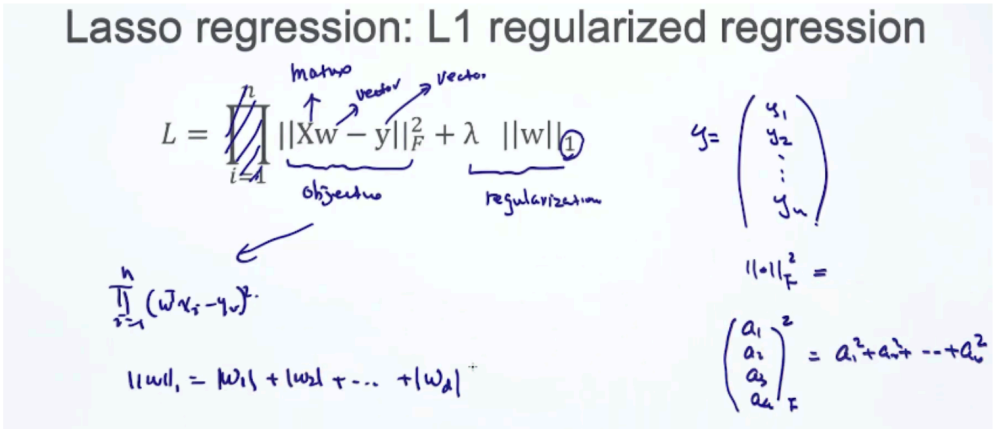


Option3: via Regularization - A Principled Way (Lasso)

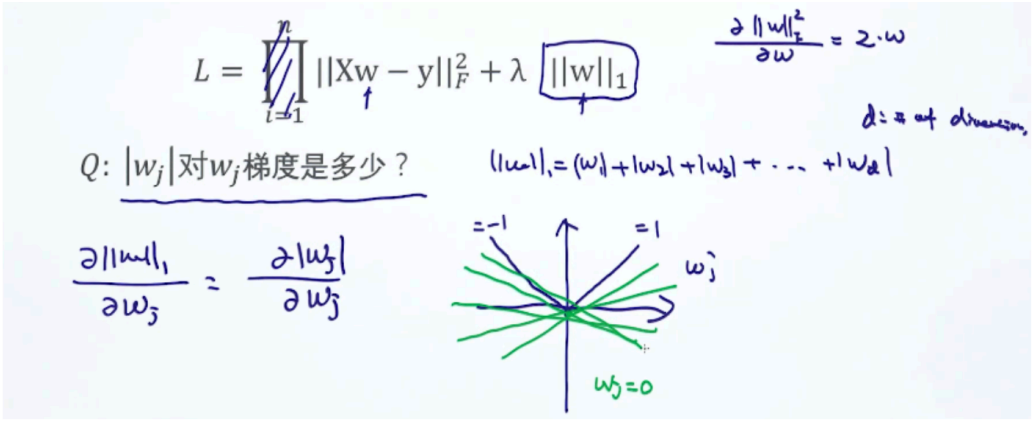
1.2 Lasso regression



可以避免 w 变得很大。但是不会让 w 变零，也就是不会稀疏。
不具备特征选择的功能。



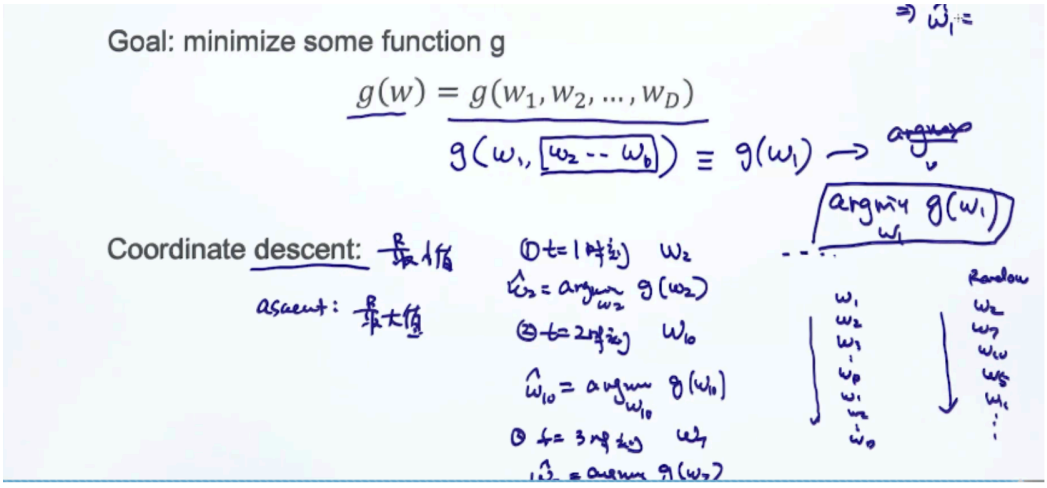
1.3 optimize Lasso objective

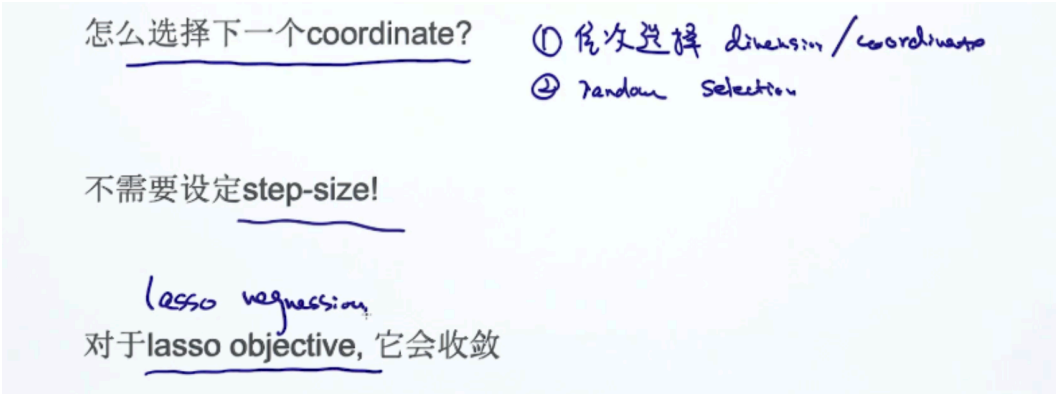


2. Coordinate Descent

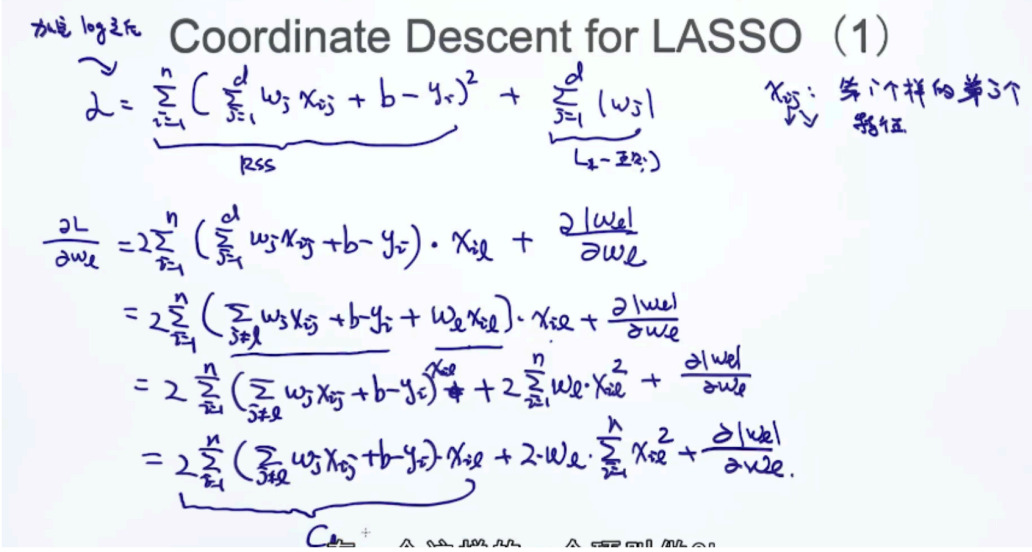
2.1

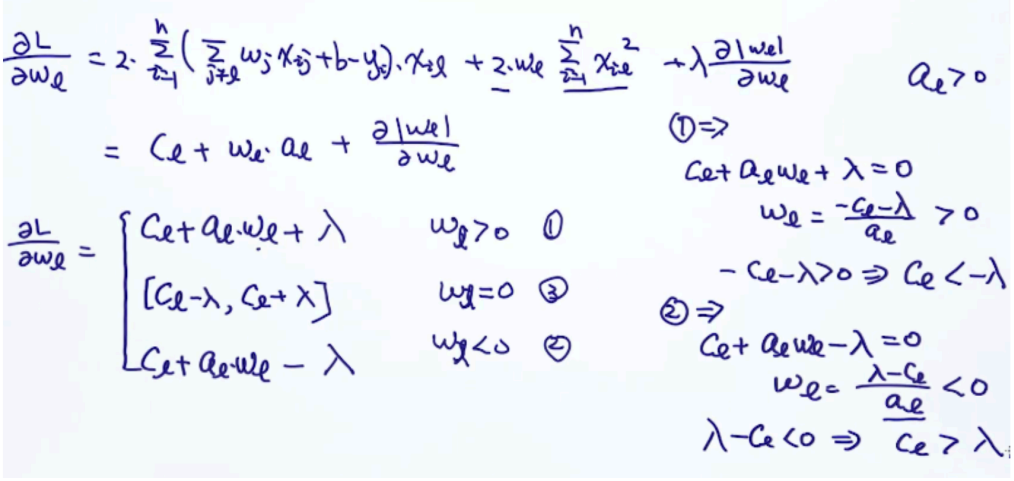
函数有很多维度，寻求最优解的时候每次只考虑一个维度。

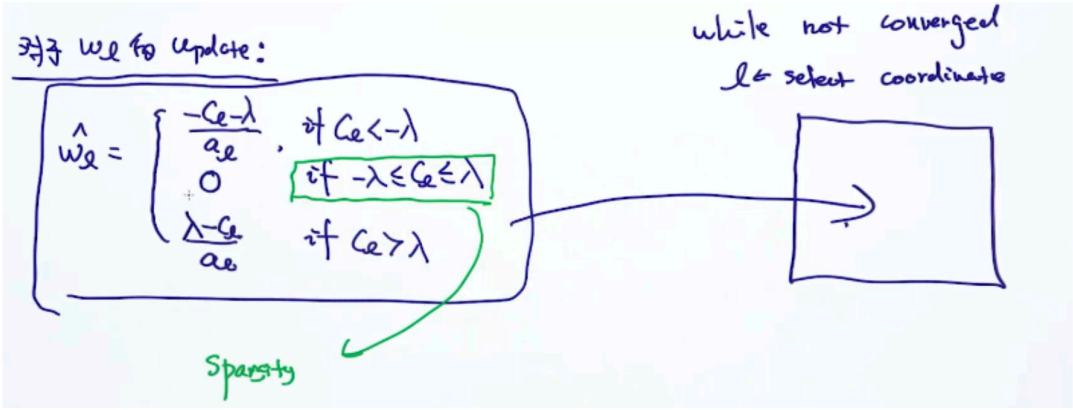




2.2 coordinate descent for Lasso







其中在更新的时候，强行将 w 设置为 0，这就是这会稀疏的原因。

2.3 other Lasso solvers

Classically: Least angle regression (LARS) [Efron et al. ‘04]

Then: Coordinate descent algorithm [Fu ‘98, Friedman, Hastie, & Tibshirani ’08]

Now:
Parallel CD (e.g., Shotgun, [Bradley et al. ‘11])
Other parallel learning approaches for linear models:
- Parallel stochastic gradient descent (SGD) (e.g., Hogwild! [Niu et al. ’11])
- Parallel independent solutions then averaging [Zhang et al. ‘12]

Alternating directions method of multipliers (ADMM) [Boyd et al. ’11]