

## Final Exam:

- Similar to midterm
- Cumulative
- 2pm-4pm (class zoom room at that time to answer questions)
- Exam posted at 12:01 am
- Accept submissions with no penalty until 11:59 pm.

Spire: Exam: May 11  
2pm-4pm

How can we ensure that a model is safe to use?

Idea: Make sure loss is small.

$$l(w_k) = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

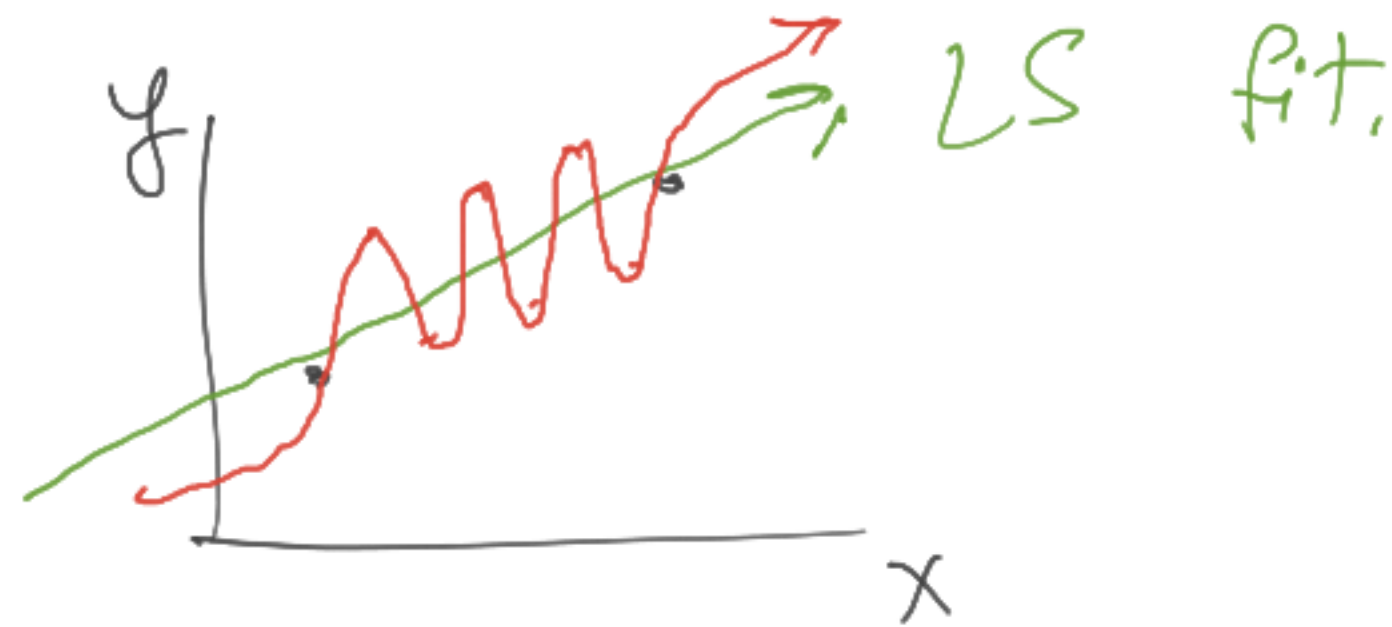
Problem: False negative worse than false pos.

Soln: Change objective to place different weights on different kinds of errors.

e.g.  $\sum_{i=1}^n \frac{10}{\substack{y_i = \text{malignant} \\ \hat{y}_i = \text{benign}}} + \frac{1}{\substack{y_i = \text{benign} \\ \hat{y}_i = \text{malignant}}}$

$$\mathbb{1}_A = \begin{cases} 1 & \text{if } A \text{ is true} \\ 0 & \text{otherwise} \end{cases}$$

Problem:



$$l(w_k) = 0.$$

- Loss  $l(w_k) = 0 \Rightarrow$  perfect model?

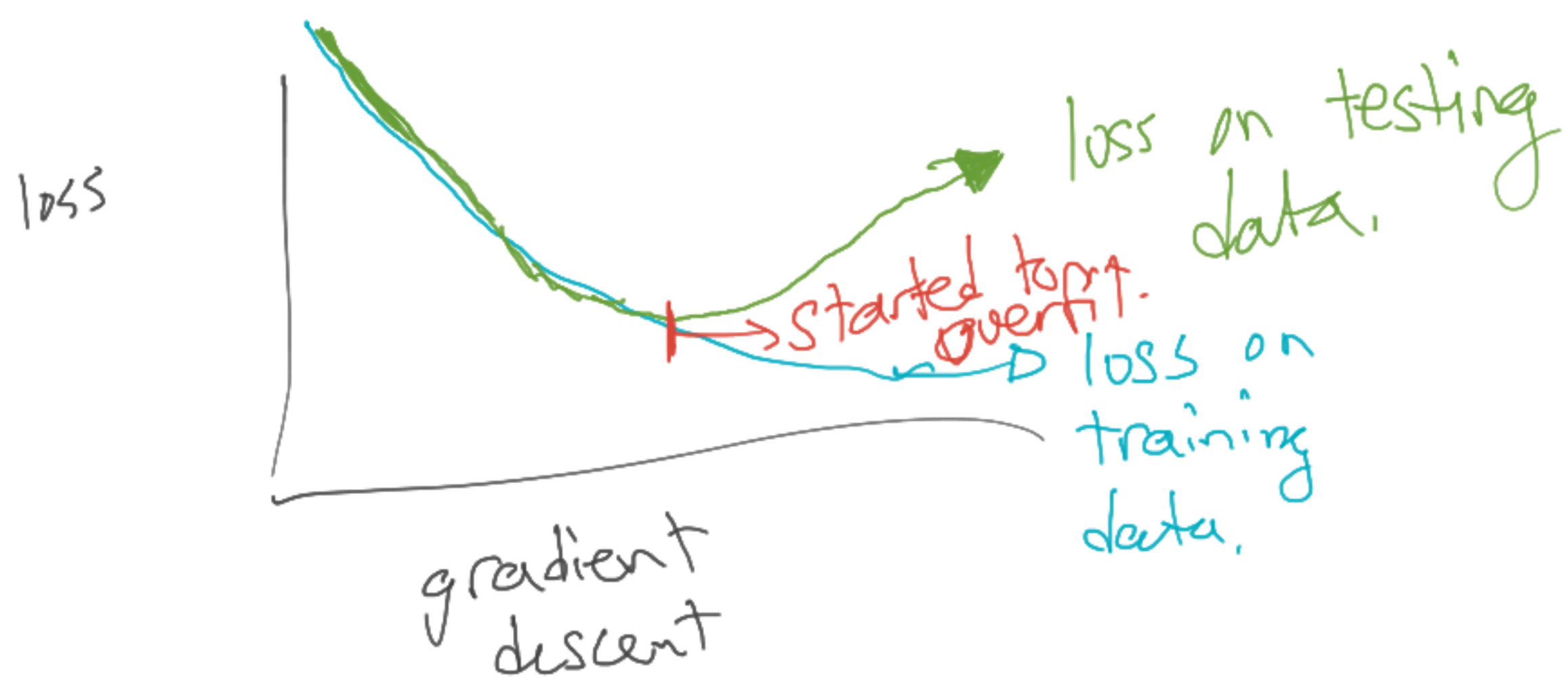
- No - "overfit" to training data.

- Issue even with lots of data.

Idea: Split data into two sets.

Training data: train model / learn weights

Test data: Used to check accuracy/performance after training.



Idea: Stop training when loss on test set increases.

Problem: You are using test data as part of model-training procedure! No longer have independent unbiased estimate of loss if model were to be used.

## Solution:

### Train

Used to train model.

### Validation

Used to set hyper-params, like when to stop training.

can be reshuffled.

- leave one-out cross validation.

### Testing

Used to test accuracy of final model.

(never used when training model)