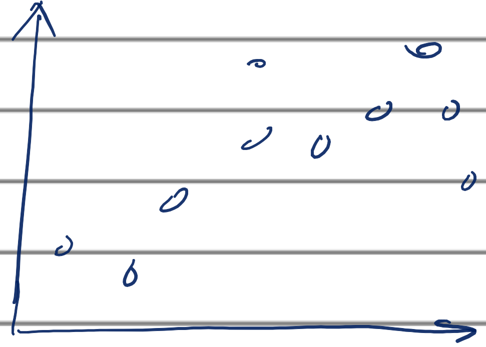
Regission:

features. covariate or predictor

Respond

line fiting

$f_w(x) = W_0 + W_1 x$



$W = (W_0, W_1)$.

fit: cost. Residual.

$R_{ss} = \sum (Y_{actual} - Y_{predict})^2$; Residual sum squ.

what if not linear: quadratic x^n. another feature.

overfiting.

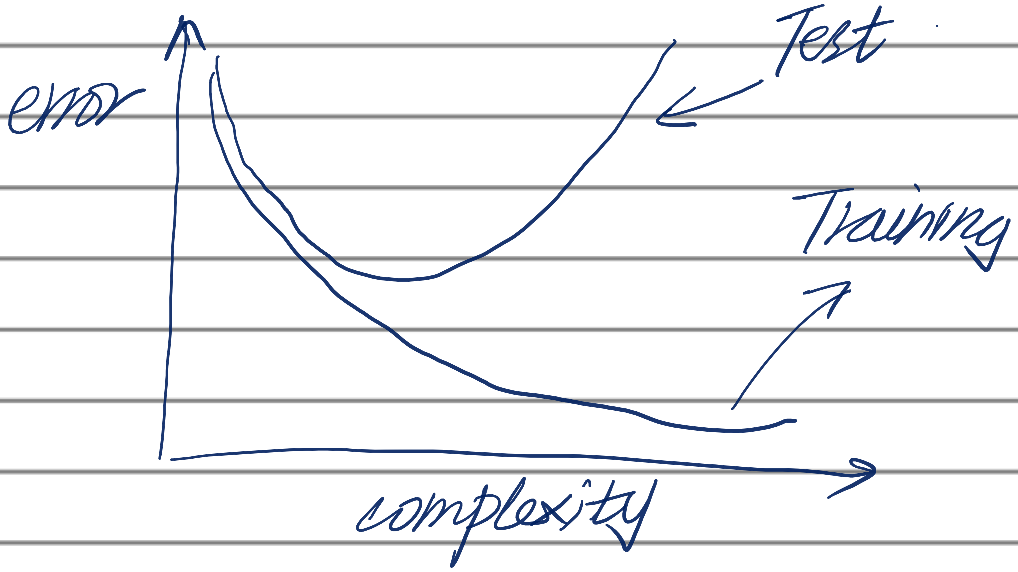
simulate prediction. remove some examples

model — remaining example ⟷ removed.

Training set ⟷ Test set.

Training error. → min → model.

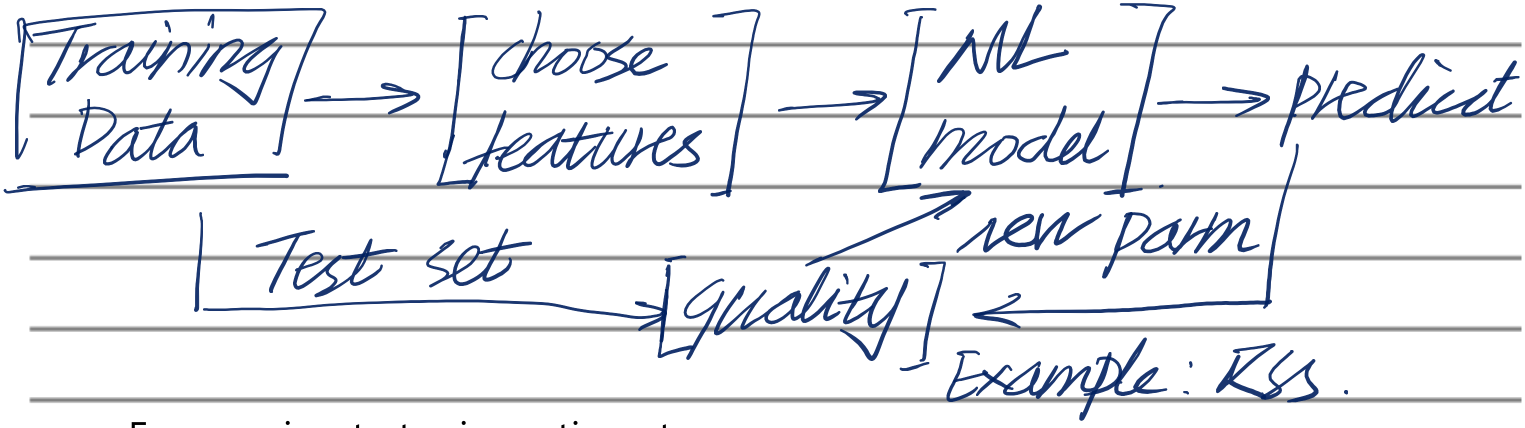
Test error ← model ⟷ Test set.



add more features.

Other Example: Salary. Stock prediction.

Data → ML → Intelligence.



From review to topic sentiments

Select a category use sentiment classifier

Classifier: Sentence from review classifier rate

Spam filter (sender, content, image content)

Medical diagnosis, brain scan (mind reading)

Linear classifier:

- Positive word, negative sentence (count the word frequency)

Sum the score

Decision boundary: separate the negative sum and positive sum

Training and evaluating classifier:

- FP and FN
- Confusion Matrix

Learning Curve: bias of model (accuracy will not be 0 with infinite samples)

Example: not good, complex model have less bias

Class probability:

- P(y|x)