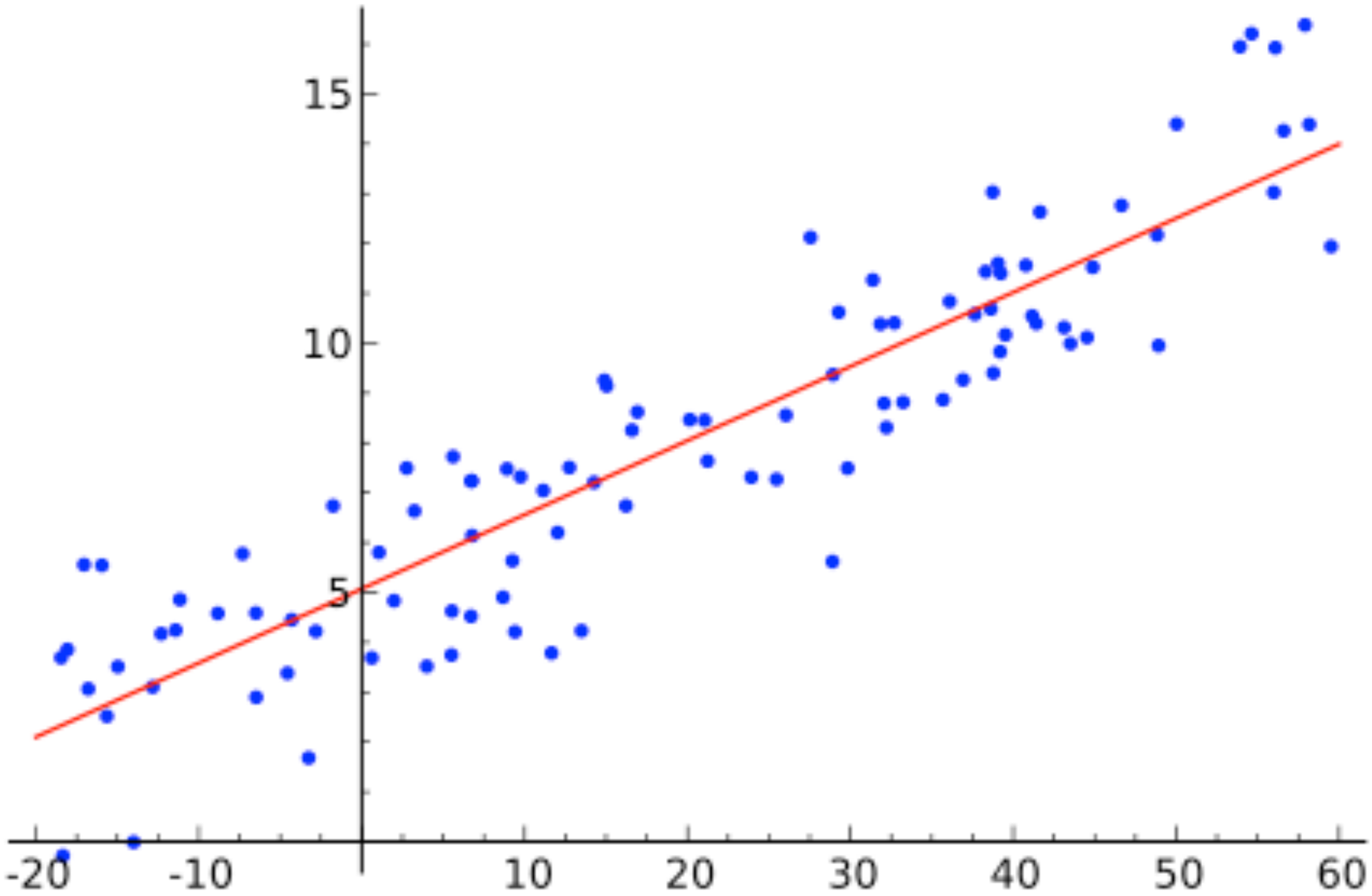
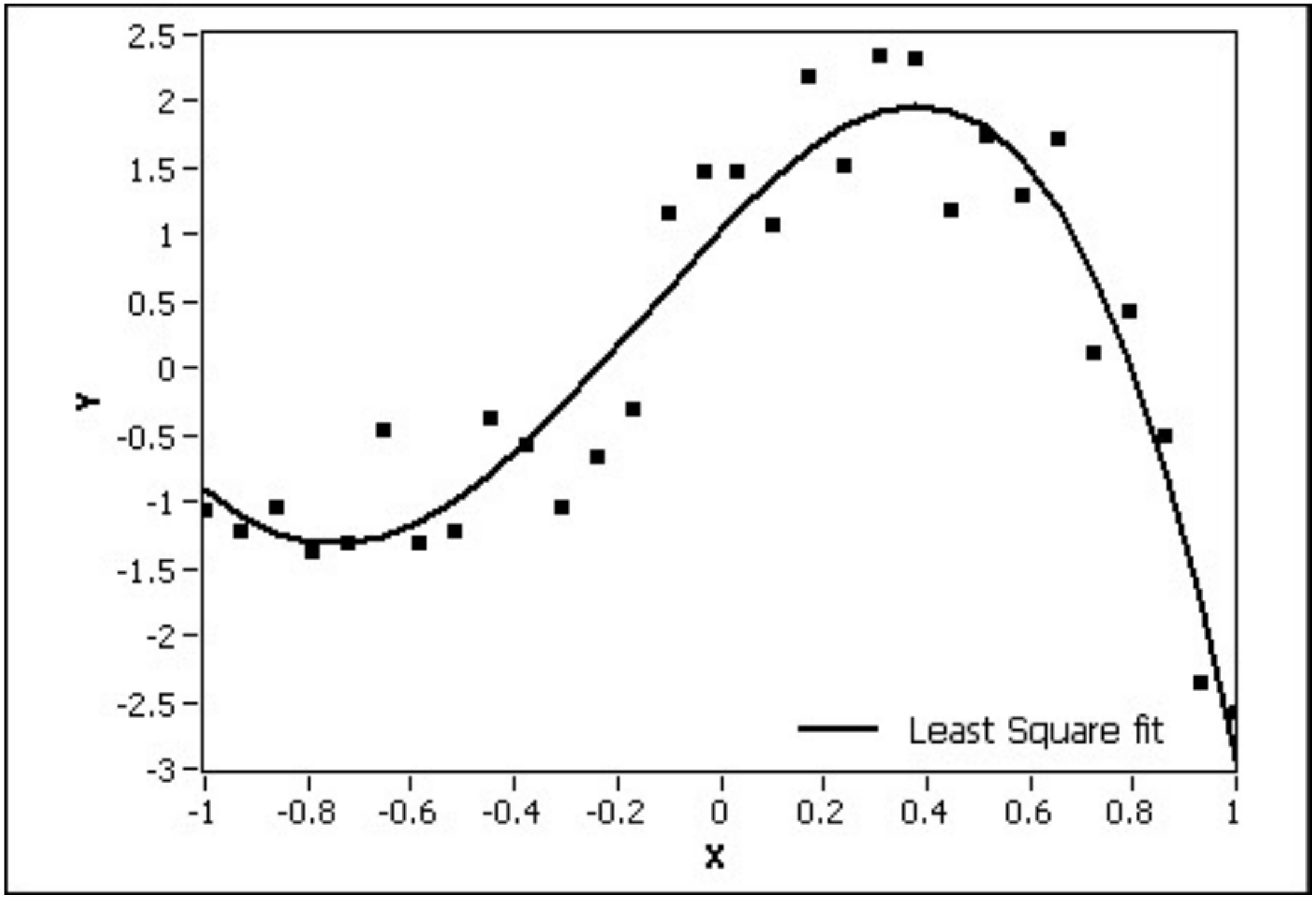
Machine Learning Basics

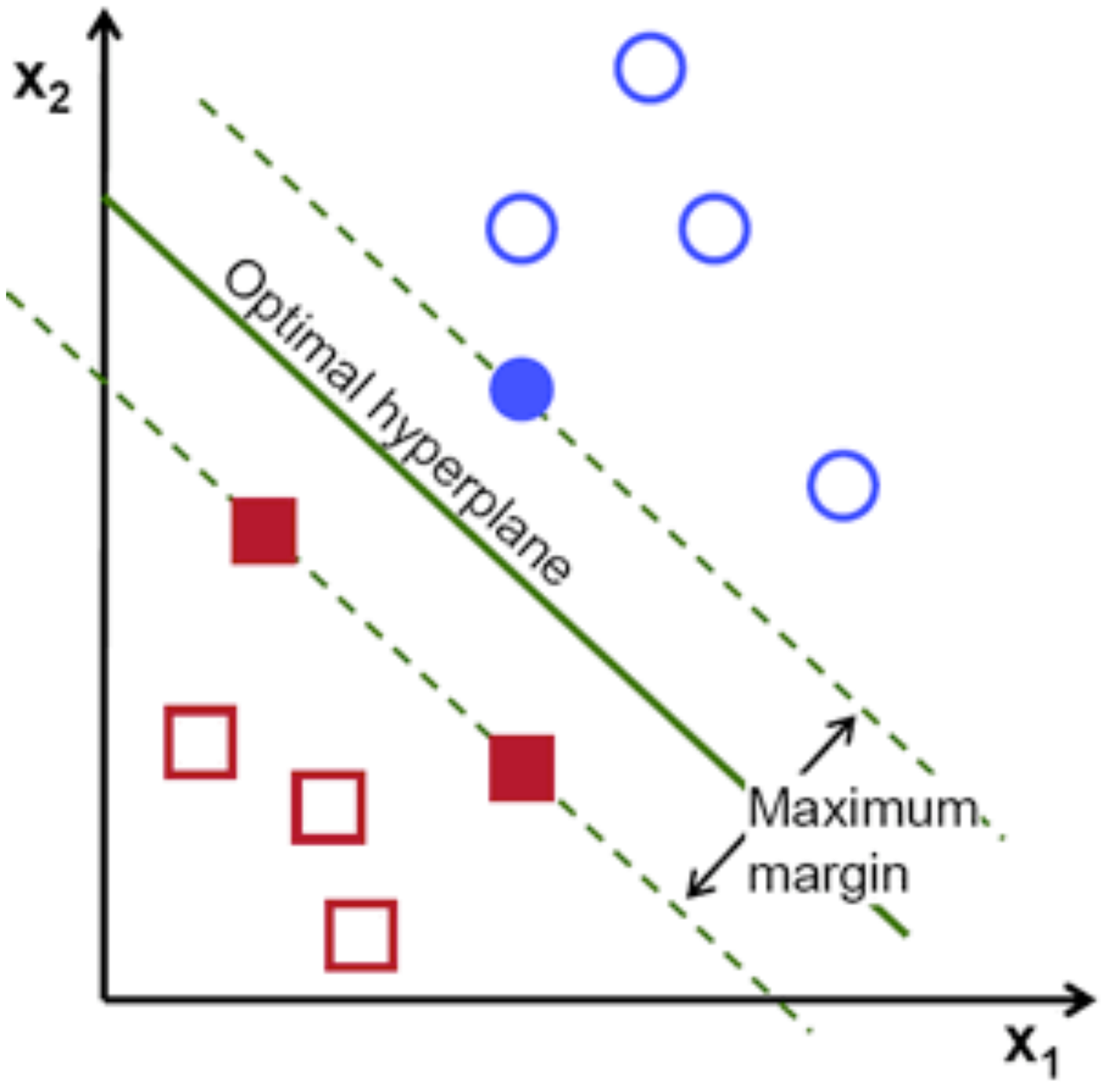
Ruihong Huang
Texas A&M University

Generative Fitting




2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
-3
Y
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
X
Least Square fit

Discriminative Fitting



Supervised Learning

- Naive Bayes assumes words are independent, but works surprising well for text classification.
- SVMs tend to work well.
- Decision trees
- Perceptron, easy to be trained.
- Logistic Regression gives you probabilities
- deep neural networks (with softmax scoring function) gives you probabilities too, can make good use of word embeddings.

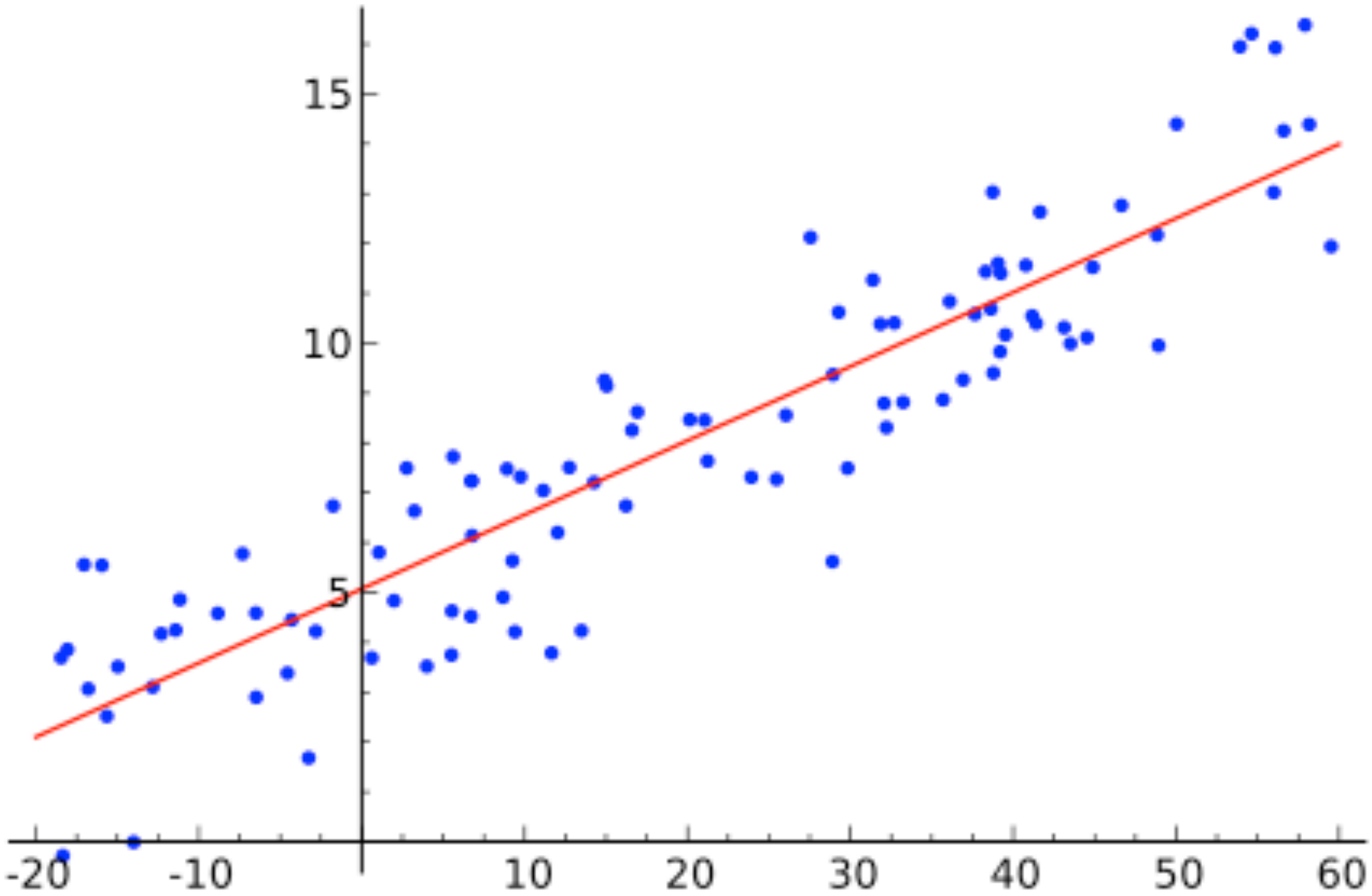
Multi-class v.s Binary

- SVMs, perceptron
- One v.s all (rest)
- One v.s one

Model Complexity & Overfitting

- model is too complex
- bad predictive performance

Generative Fitting



Tuning Data & Test Data

- Split your annotated data to training data, tuning data (also called development data) and test data.
- use tuning data to tune your system, parameters.
- Only run over your test data when your system is finalized.

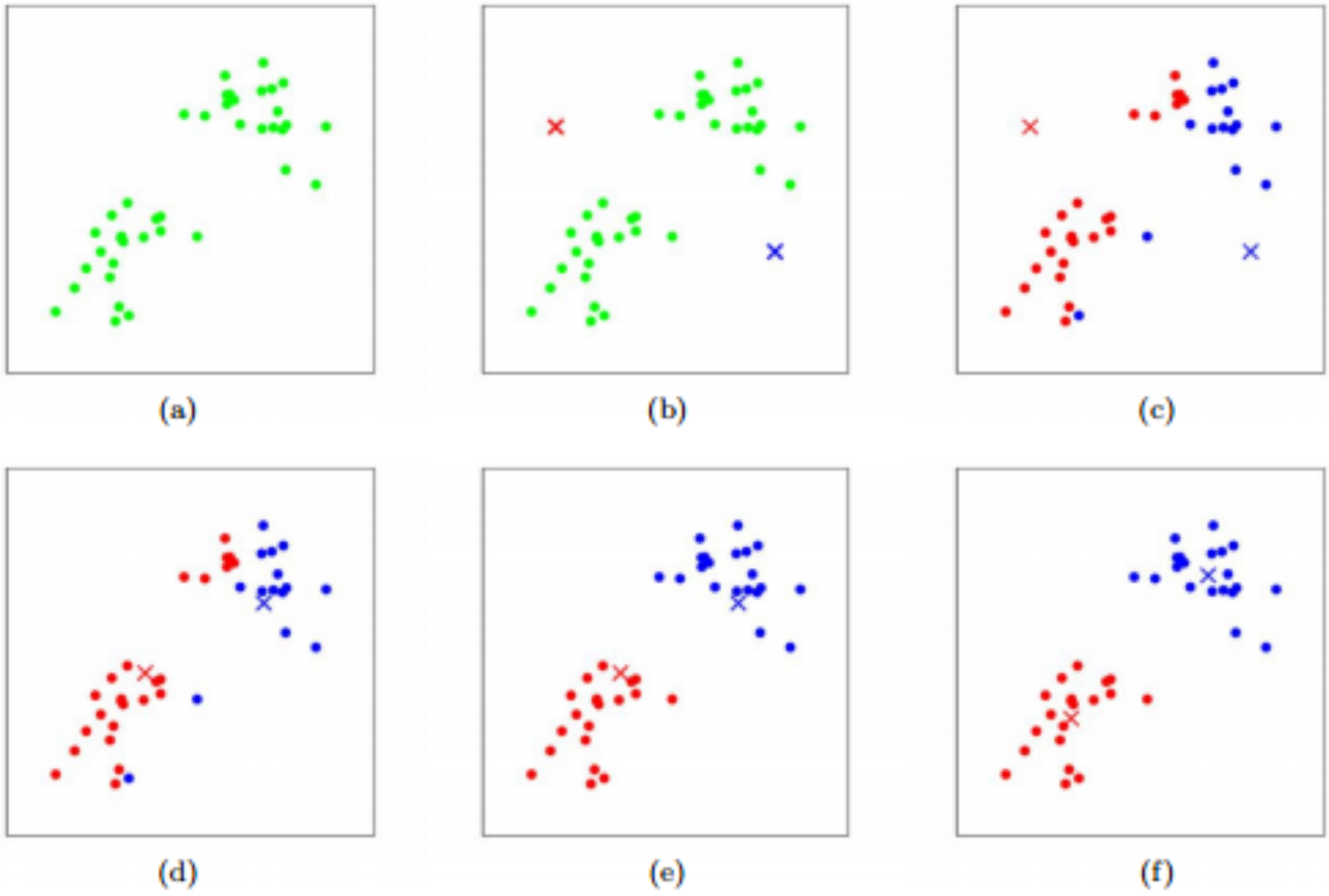
Evaluation Metric

- task dependent.
- IE tasks or similar, precision/recall/f1
- summarization: rouge
- ...

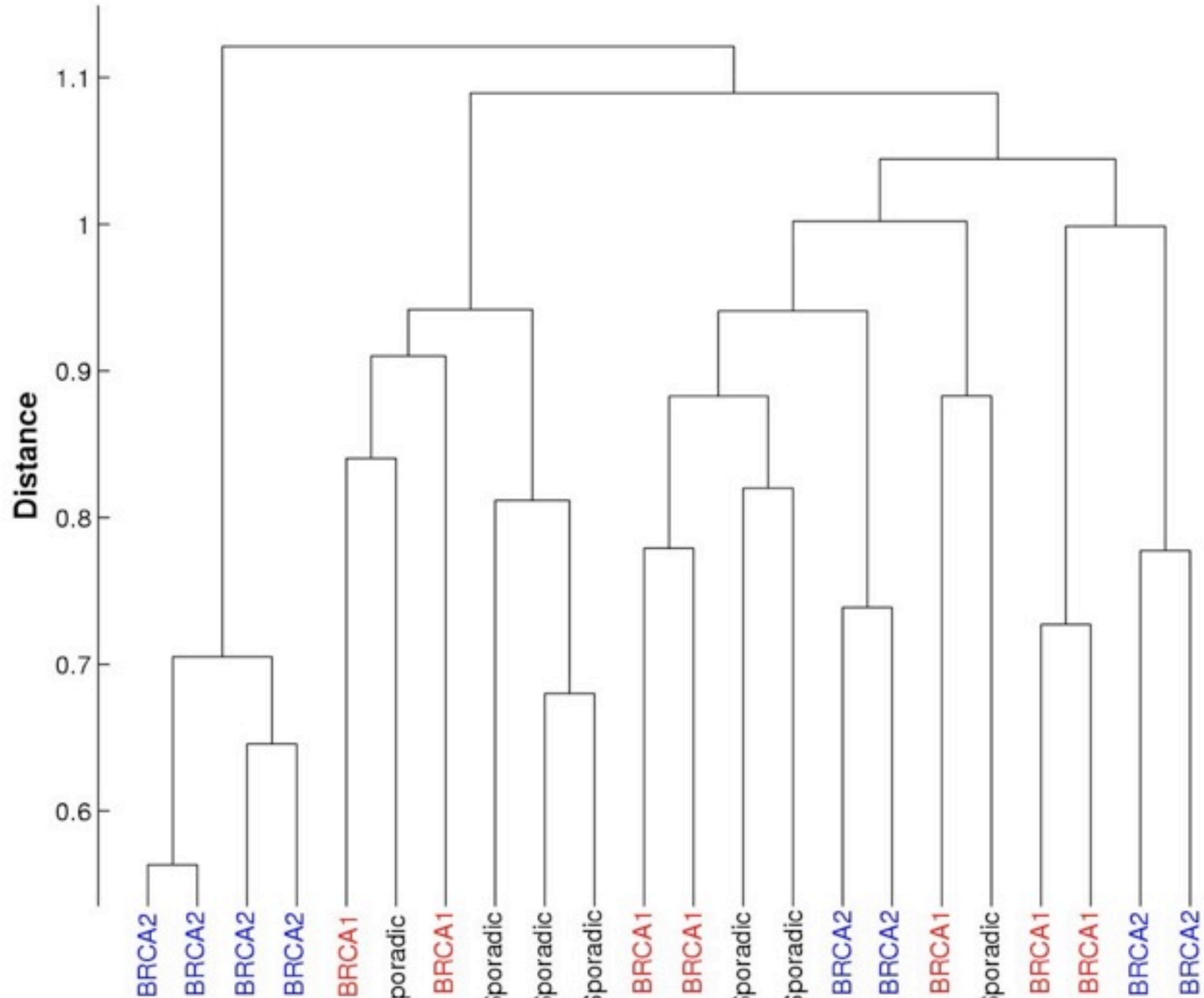
Clustering

- Top-down v.s bottom up
- Top-down example: k-means
- bottom up example: hierarchical clustering

K-means



Hierarchical Clustering



Distance Metric

- very important
- task dependent

Other learning paradigms

- Semi-supervised Learning
- active learning
- co-training
- self-training