Ask questions or say hi in chat before/during/after class!

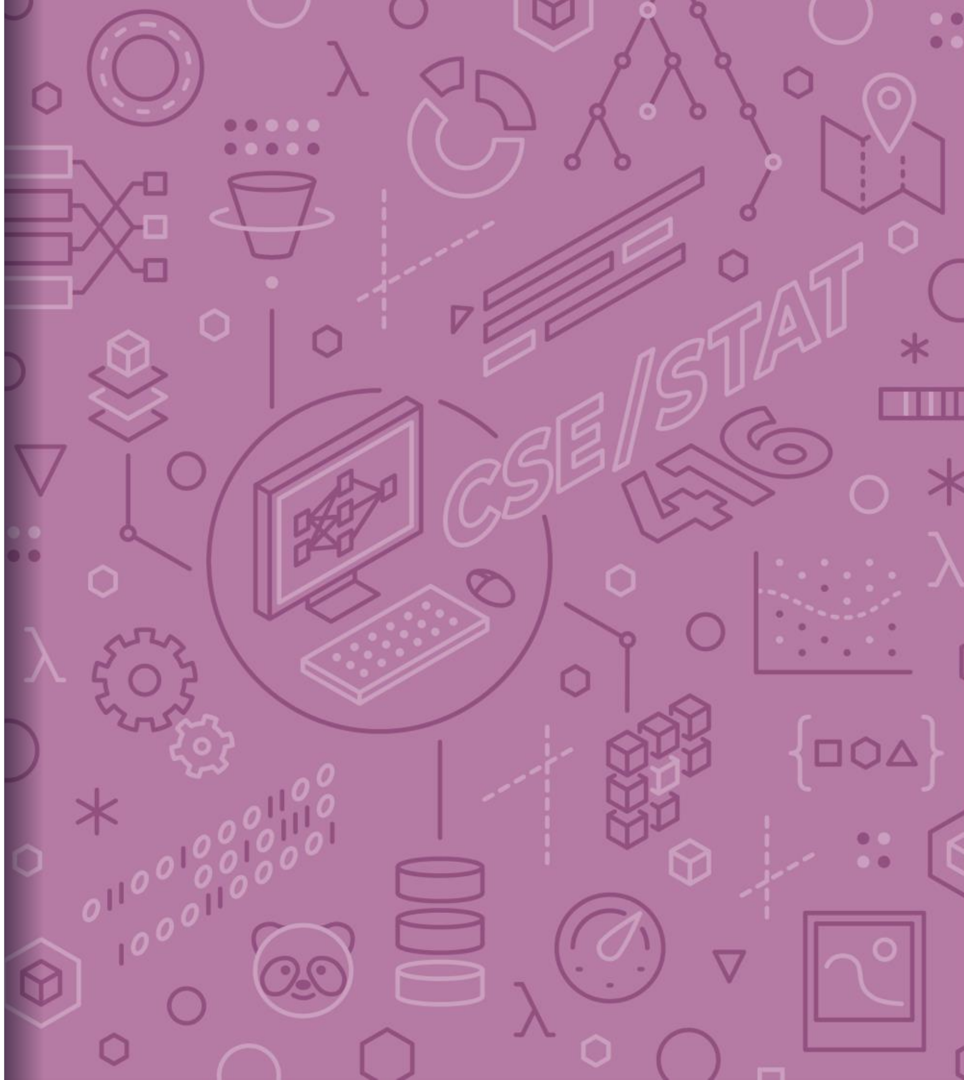
CSE/STAT 416

Course Wrap Up

Hunter Schafer
Paul G. Allen School of Computer Science & Engineering
University of Washington

June 2, 2021

Music: Manatee Commune



Administrivia

- Final Exam
  - Monday 8:30 am – Wednesday 8:30 am
  - Open note and open internet, but don't have someone else solve problems for you
  - Can be taken as a group (submit together on Gradescope)
  - No OH or help with review/conceptual questions during the exam. Help available on EdStem for clarification or logistic questions only.
  - Extra OH this weekend to help review
- Learning Reflection: Reminder that this week's learning reflection is a mind map instead of the usual format!
- One last checkpoint for today's class (if necessary), due on Monday like our normal schedule
- **Please fill out the course evals!**

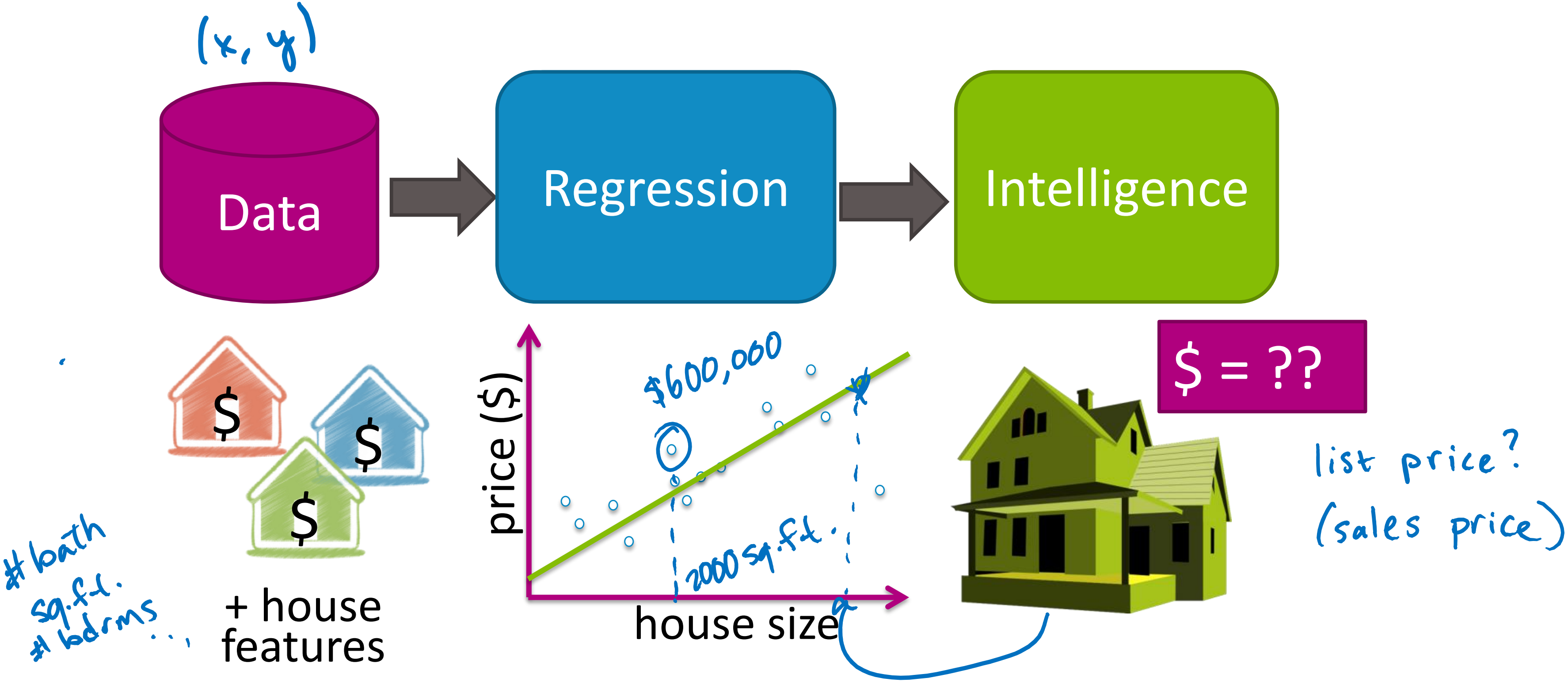
One Slide

- Regression
- Overfitting
- Training, test, and generalization error
- Bias-Variance tradeoff
- Ridge, LASSO
- Cross validation
- Gradient descent
- Classification
- Logistic regression
- Bias / Fairness
- Decision trees
- Boosting
- Precision and recall
- Nearest-neighbor retrieval, regression, and classification
- Kernel regression
- Locality sensitive hashing
- Dimensionality reduction, PCA
- k-means clustering
- Hierarchical clustering
- Unsupervised v. supervised learning
- Recommender systems
- Matrix factorization
- Coordinate descent
- Neural networks
- Convolutional neural networks
- Transfer learning for deep learning

Case Study 1:
Predicting house prices

Model: $y_i = f(x_i) + \epsilon_i$

Predictor: $\hat{y}_i = \hat{f}(x_i)$



Regression

Ridge: $\arg\min_w L(w) + \lambda \|w\|_2^2$

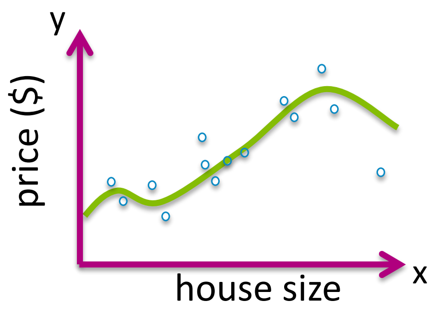
Case study: Predicting house prices

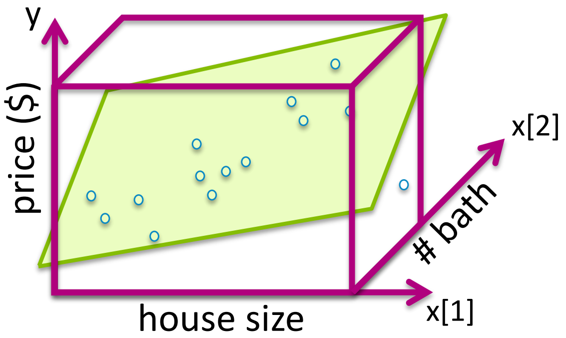
Models

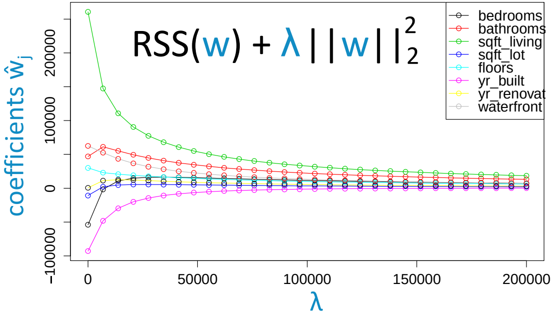
- Linear regression
- Regularization: Ridge (L2), Lasso (L1)

Including many features:

- Square feet
- # bathrooms
- # bedrooms
- Lot size
- Year built
- …







Regression

Case study: Predicting house prices

Quality Metric: $RSS(w) = \sum_{i=1}^{n} (w^T h(x_i) - y_i)^2$

$\hat{w} = \underset{w}{\text{argmin}}\ RSS(w)$

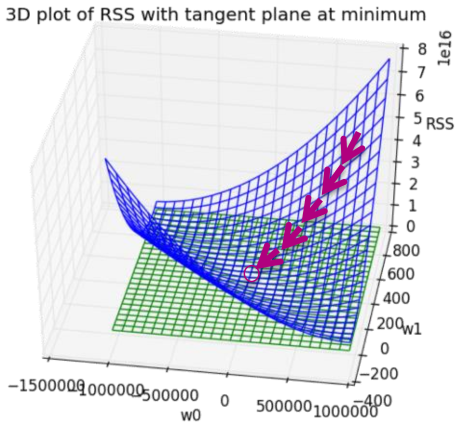
Algorithms

- Gradient descent

$RSS(w_0,w_1) =$
$(\$_{house\ 1}-[w_0+w_1 sq.ft._{house\ 1}])^2$
$+ (\$_{house\ 2}-[w_0+w_1 sq.ft._{house\ 2}])^2 +$
$(\$_{house\ 3}-[w_0+w_1 sq.ft._{house\ 3}])^2 + \ldots$
[include all houses]

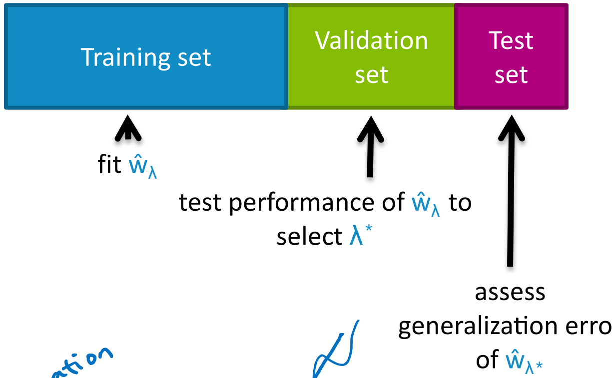
$\downarrow$

$\hat{w}$



Regression

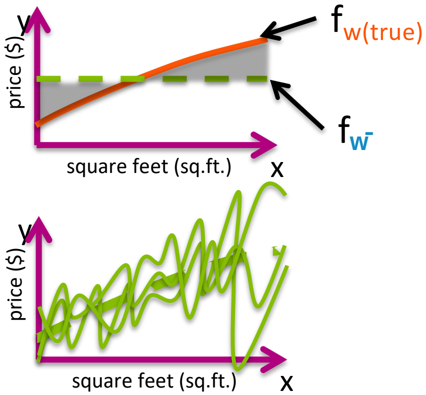
Case study: Predicting house prices

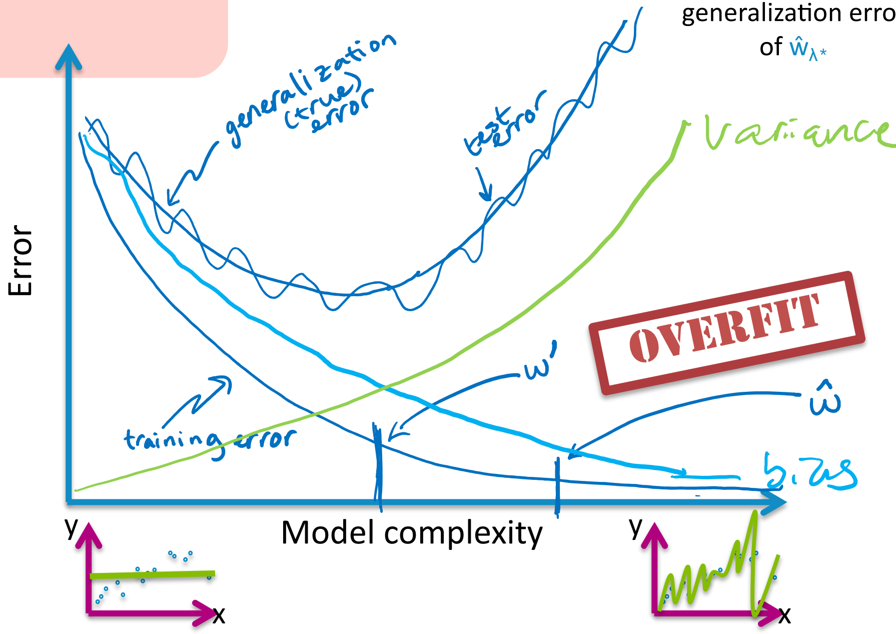


Concepts

- Loss functions, bias-variance tradeoff, cross-validation, sparsity, overfitting, model selection

1. Noise
2. Bias
3. Variance





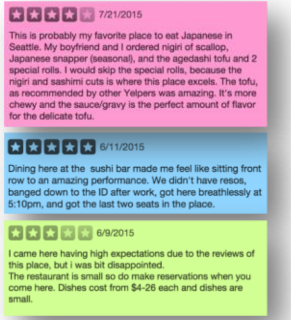
Case Study 2:
Sentiment analysis

Sushi was awesome,
the food was awesome,
but the service was awful.

stars / +/-

text

All reviews:



7/21/2015

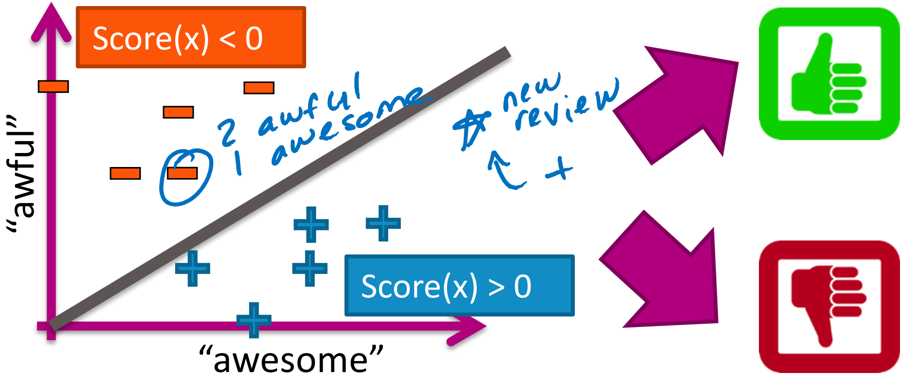
This is probably my favorite place to eat Japanese in Seattle. My boyfriend and I ordered nigiri of scallop, Japanese snapper (seasonal), and the agedashi tofu and 2 special rolls. I would skip the special rolls, because the nigiri and sashimi cuts is where this place excels. The tofu, as recommended by other Yelpers was amazing. It's more chewy and the sauce/gravy is the perfect amount of flavor for the delicate tofu.

6/11/2015

Dining here at the sushi bar made me feel like sitting front row to an amazing performance. We didn't have resos, banged down to the ID after work, got here breathlessly at 5:10pm, and got the last two seats in the place.

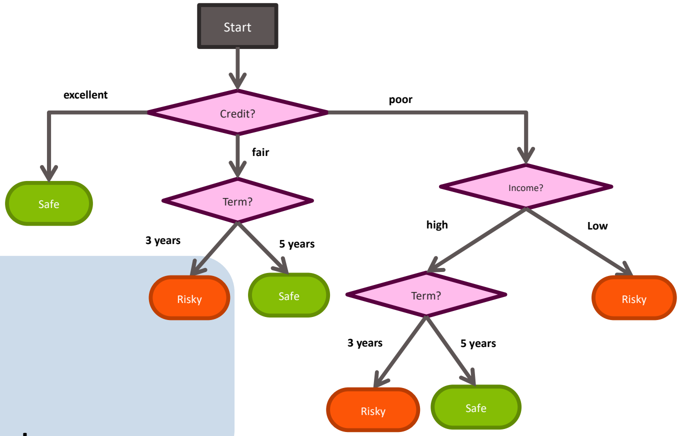
6/9/2015

I came here having high expectations due to the reviews of this place, but i was bit disappointed.
The restaurant is small so do make reservations when you come here. Dishes cost from $4-26 each and dishes are small.



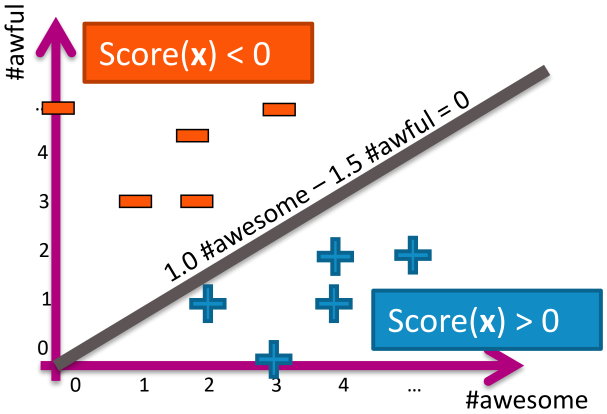
Classification

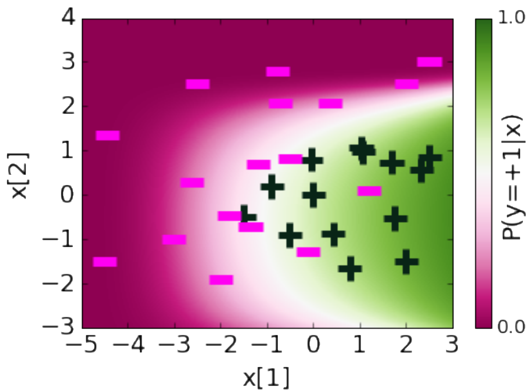
Case study: Analyzing sentiment



Models

- Linear classifiers (logistic regression)
- Multiclass classifiers
- Decision trees
- Boosted decision trees and random forests





AdaBoost

$$\hat{f}(x_i) = \text{sign}\left(\sum_{t=1}^{T} \hat{w}_t \hat{f}_t(x_i)\right)$$

↑ model coefs.

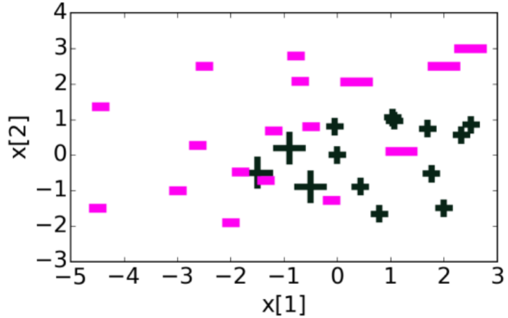
α_i ← dataset weights

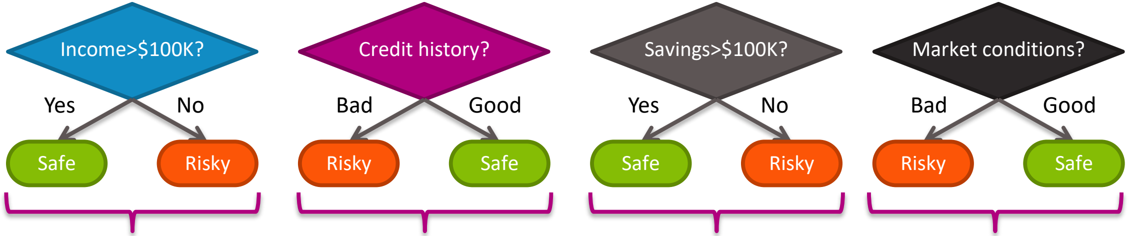
Classification

Case study: Analyzing sentiment

Algorithms

- Boosting
- Learning from weighted data α_i





weighted_error = 0.2

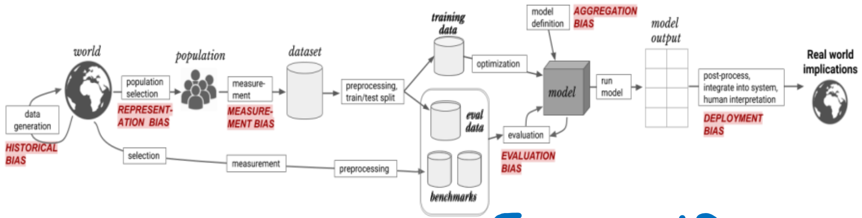
weighted_error = 0.35

weighted_error = 0.3

weighted_error = 0.4

Classification

Case study: Analyzing sentiment

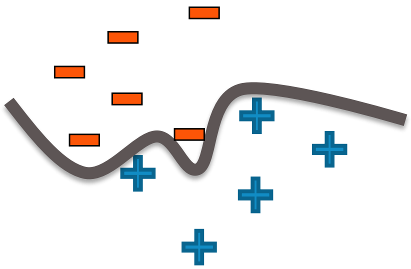


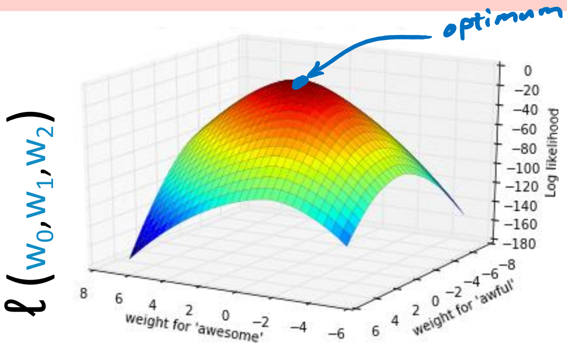
Fairness / Bias

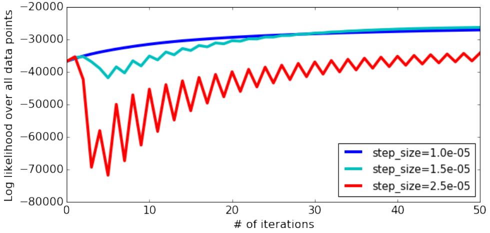
Concepts

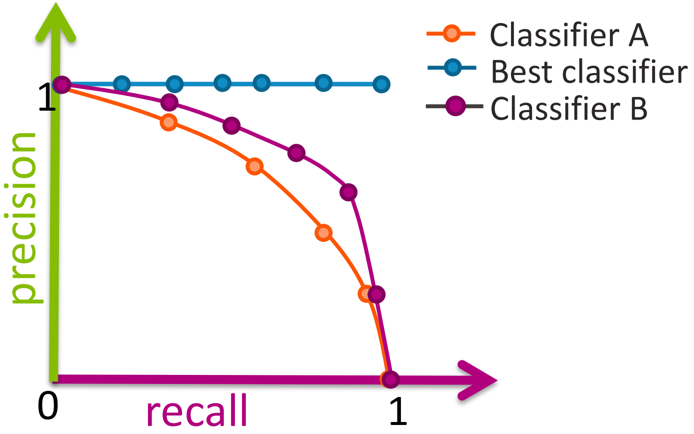
- Decision boundaries, maximum likelihood estimation, ensemble methods, random forests
- Precision and recall

Accuracy isn't always enough
↳ class imbalance
↳ unfair treatment









Case Study 3:
Document retrieval

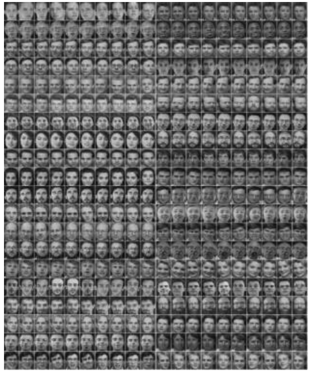
- embedding (BoW vs. TF-IDF)
- distance metrics (euclidean, manhattan, cosine)

Case Study 3+:
Document structuring for retrieval

Case Study 3++:
Dimensionality reduction

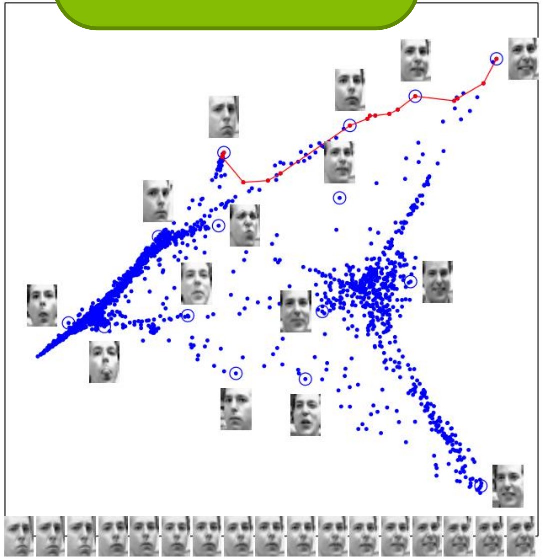
Data → PCA → Intelligence

simple



Images with thousands or millions of pixels

Can we give each image a coordinate, such that similar images are near each other?



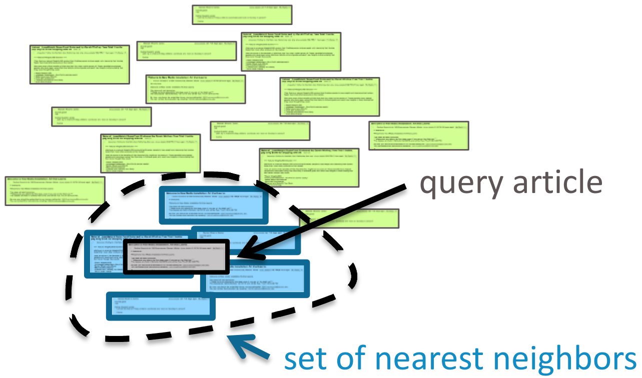
[Saul & Roweis '03]

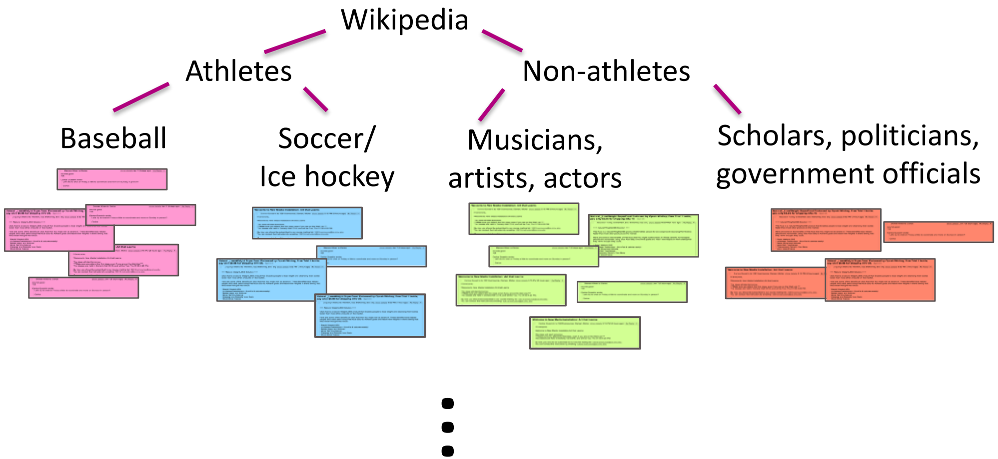
Clustering & Retrieval

Case study: Finding documents

Models

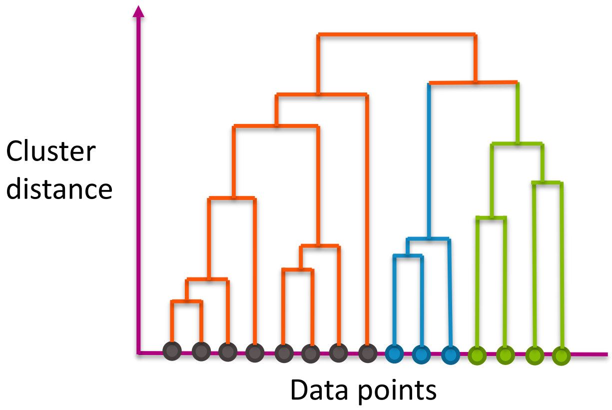
- Nearest neighbors
- Clustering
- Hierarchical clustering





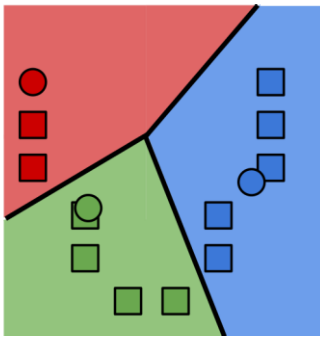
Clustering & Retrieval

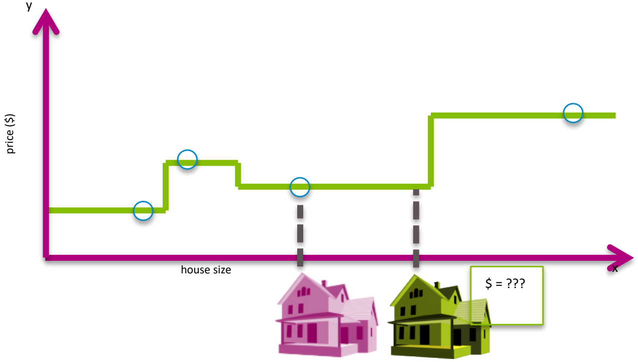
Case study: Finding documents

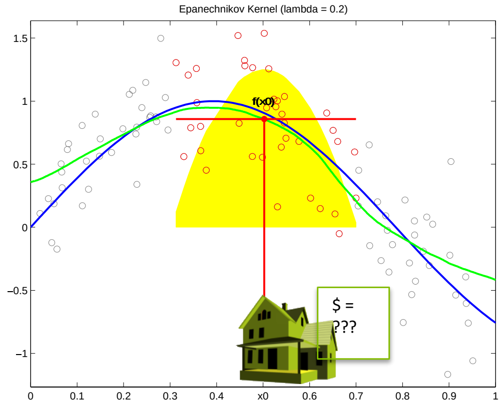


Algorithms

- k-means / k-means++
- Locality-sensitive hashing (LSH)
- NN regression and classification k-NN
- Kernel regression (kernel)
- Agglomerative and divisive clustering
- PCA





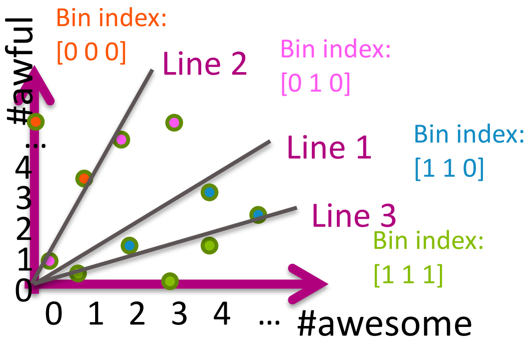


Clustering & Retrieval

Case study: Finding documents

Concepts

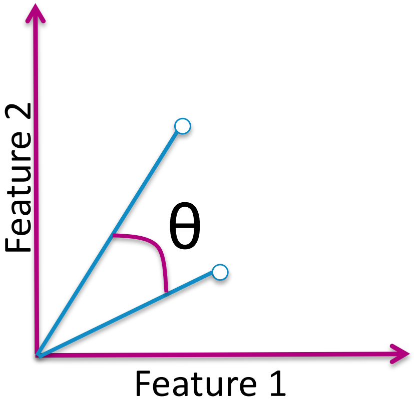
- Distance metrics, kernels, approximation algorithms, dimensionality reduction



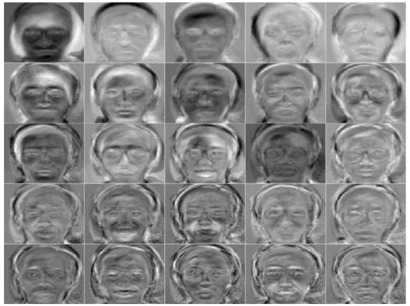
1	0	0	0	5	3	0	0	1	0	0	0	0

3	0	0	0	2	0	0	1	0	1	0	0	0

1*3
\+
5*2
= 13



Principal components:



Reconstructing:

Case Study 4: Product recommendation

Data → Matrix Factorization → Intelligence

Your past purchases:

+ purchase histories of all customers

Customers

features	
features	
features	

Products

features	
features	
features	

Recommended items:

Recommender Systems & Matrix Factorization

Case study: Recommending Products

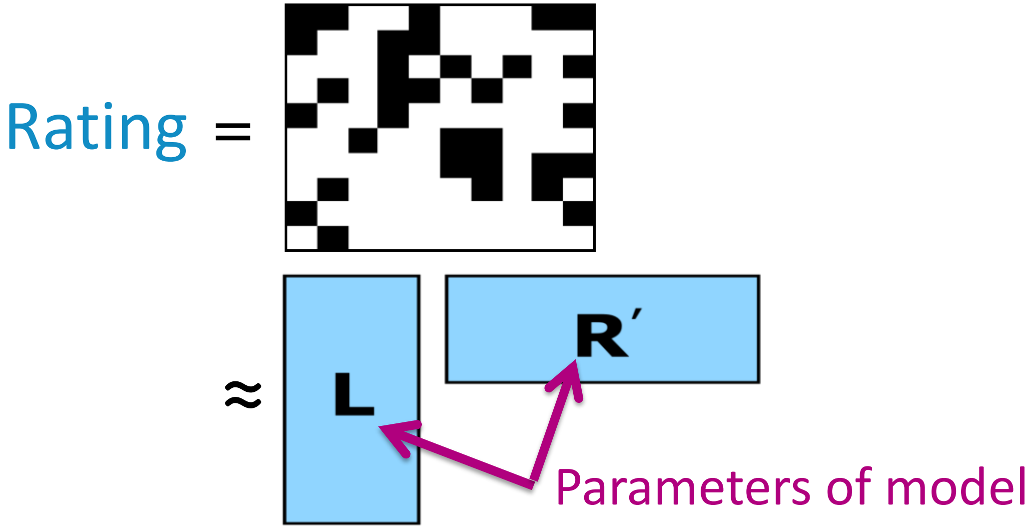
Models

- Collaborative filtering
- Matrix factorization

Popularity

Co-occurence matrix

Featurized MF

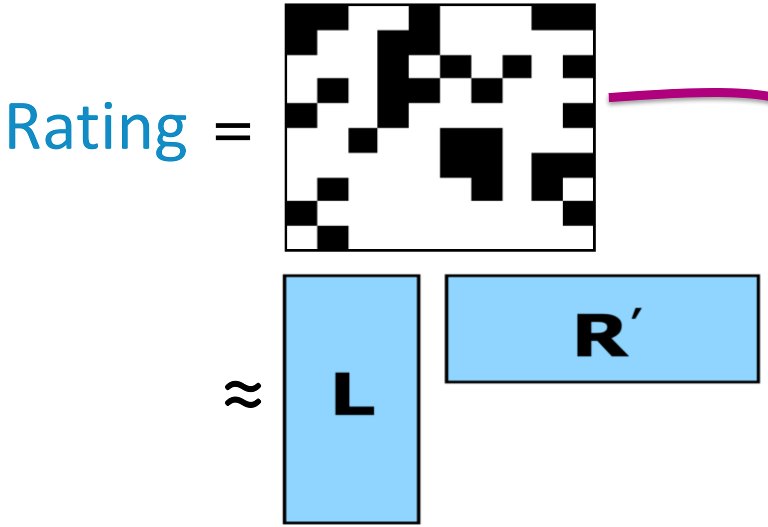


Recommender Systems & Matrix Factorization

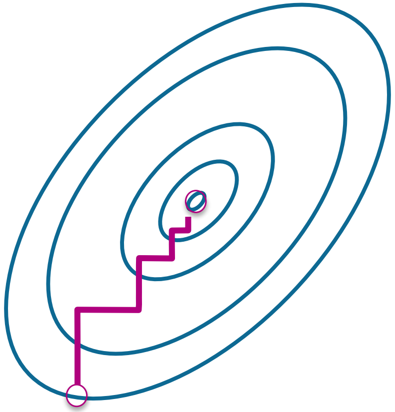
Case study: Recommending Products

Algorithms

- Coordinate descent



Form estimates $\hat{L}_u$ and $\hat{R}_v$

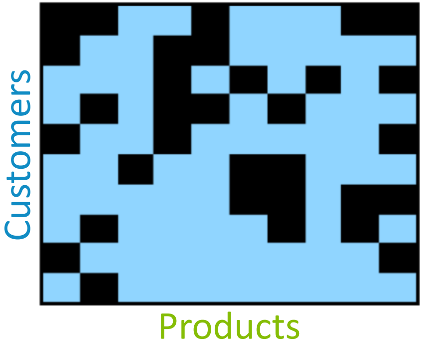


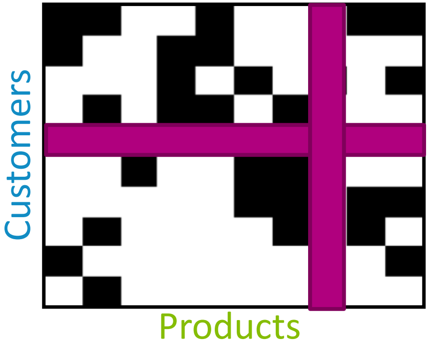
Recommender Systems & Matrix Factorization

Case study: Recommending Products

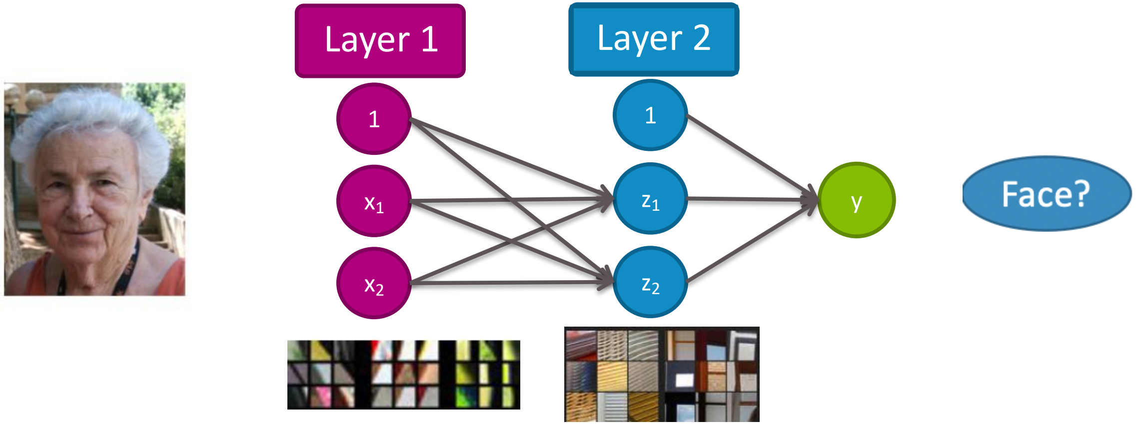
Concepts

- Matrix completion, cold-start problem





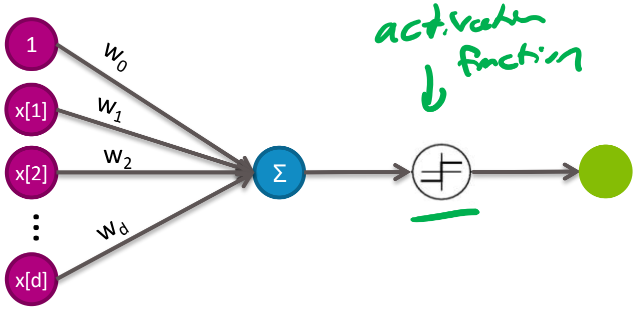
Case Study 5:
Image classification



Deep Learning

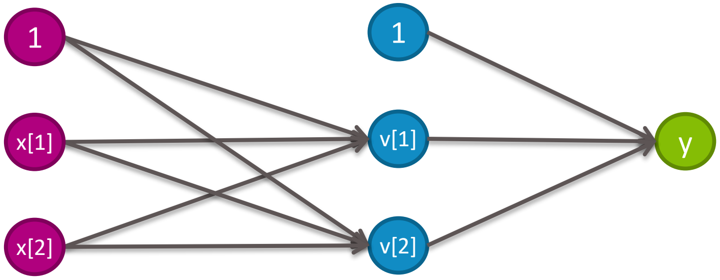
Case study: Image classification

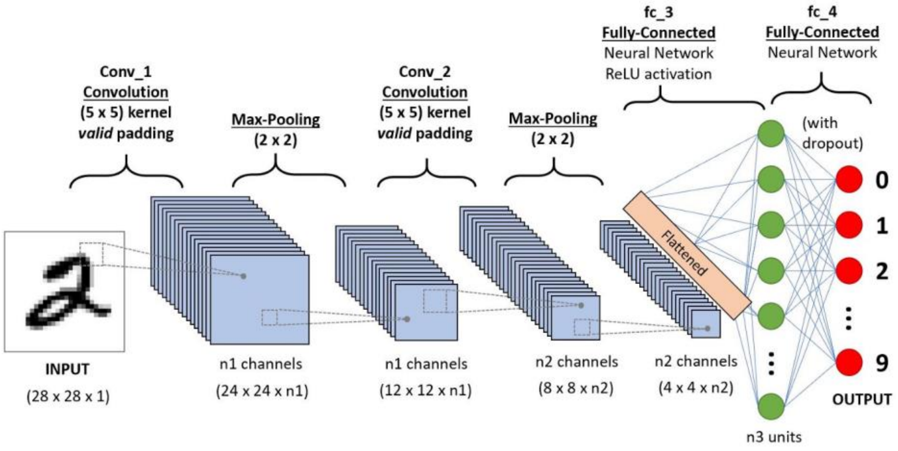
Fully connected NN



Models

- Perceptron
- General neural network
- Convolutional neural network



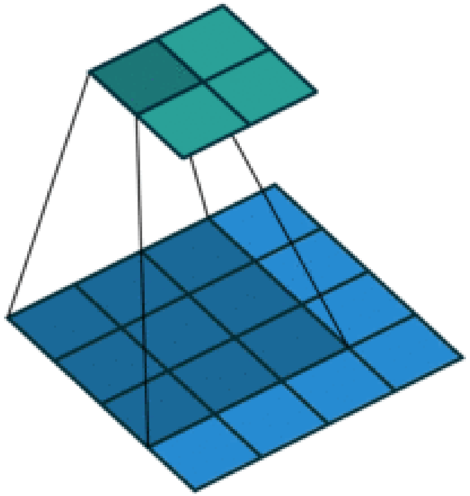


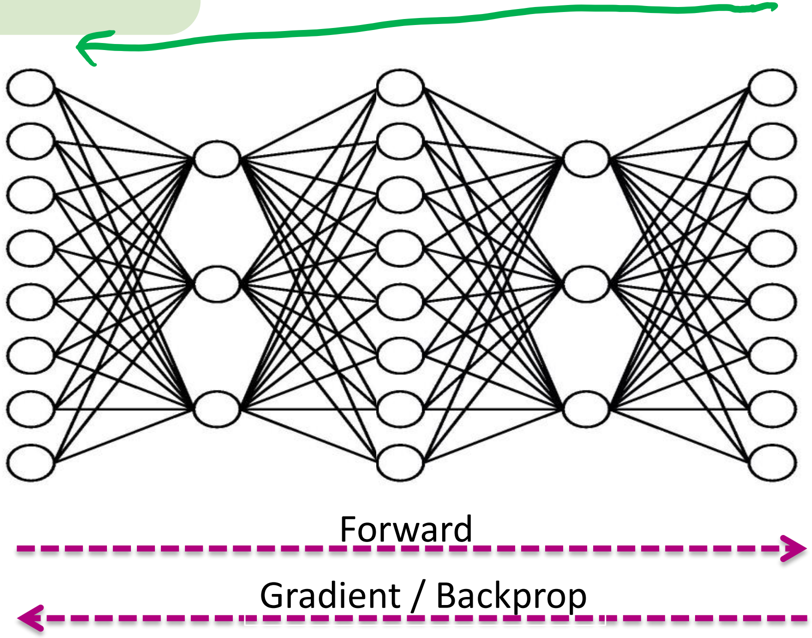
Deep Learning

Case study: Image classification

Algorithms

- Convolutions
- Backpropagation (high level only)



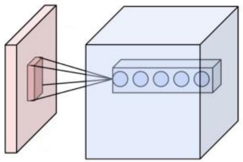


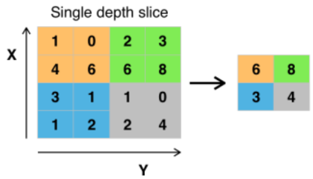
Deep Learning

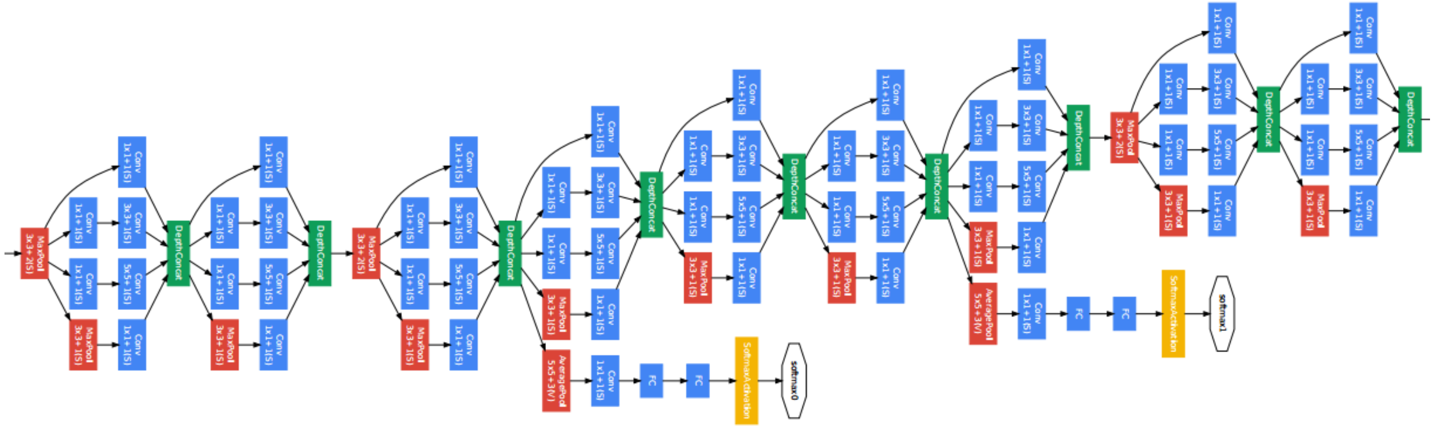
Case study: Image classification

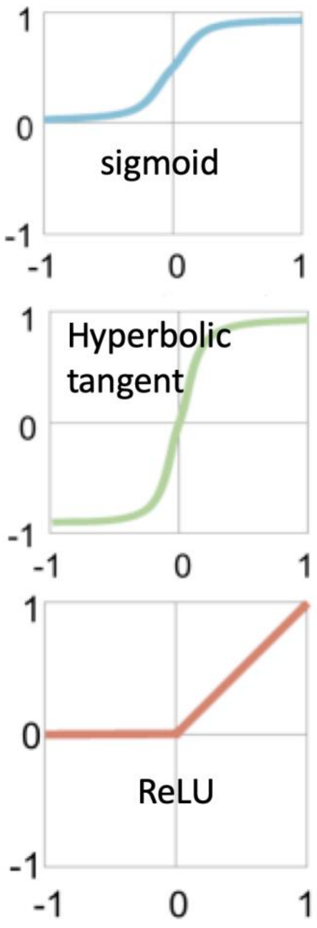
Concepts

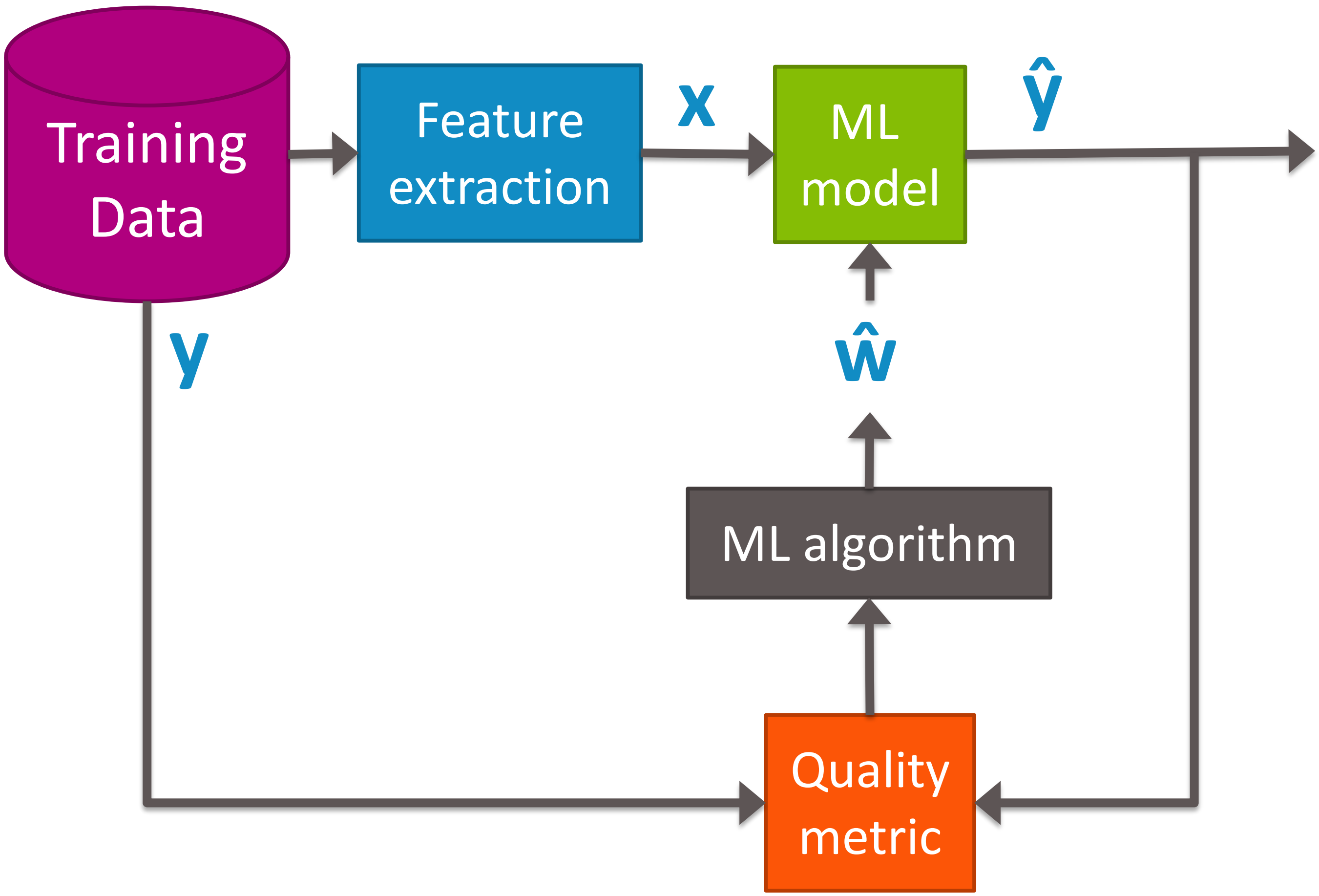
- Activation functions, hidden layers, architecture choices










Training Data
Feature extraction
x
ML model
ŷ
y
ŵ
ML algorithm
Quality metric

3:35

Brain Break

Poll Everywhere

Think

2 min

pollev.com/cs416

Tomorrow's quiz section will be Q&A review sessions. You bring questions and your TA will review concepts and discuss with your section.

To help them do a little prep please fill out this PollEverwhere question outlining some topics or questions you would want them to go over in section.

Free response box, but please put your section at the beginning (e.g., "Section AF: I would want to go over matrix factorization").

This poll will be open all day today, but will close tonight.

Future Directions

Classes

There isn't a clear, "one right class" to take next! If you want to take course work, you can take anything that you are interested in to apply your ML knowledge there!

Fairly comprehensive list of data science class at UW: https://escience.washington.edu/data-science-courses-at-the-university-of-washington/

Future Directions

This is a (very insufficient) attempt to outline some interesting directions ML research is going. This list fails to provide breadth of coverage and depth of all the ways ML can be applied.

Something not showing up in this list doesn't mean ML can't be used for that task! I'm just one opinion about what I'm excited about in ML!

FAccT

How do we make ML systems that don’t cause harm when interacting with complicated, human systems.

ACM Conference on Fairness, Accountability, and Transparency

- Fairness: How to define and ensure fairness
- Accountability: Law and policy, metrics and audits
- Transparency: Interpretable and explainable models
- Privacy and Security: Privacy-preserving models, federated learning
- Human-ML Interaction: Humans in the loop, UX design, community designed systems, education

Interactive Learning / Reinforcement Learning

How do we design models that interact with the environment?
Examples:

- Self driving cars and robotics
- Game agents

Areas of study:

- Interactive Learning: Multi-armed bandits
- Reinforcement Learning: Q-learning
  - Deep reinforcement learning
- Ensuring safety in interactive systems

ML Systems

Construed broadly, trying to build systems to efficiently implement ML models.

- Hardware: TPU (Google)
- Energy Efficiency:
  - Green AI
  - TinyML
- Distributed Systems: Cloud software

ML Theory

Building foundational understanding for why/how ML works.

- Learning Theory (sample complexity)
  - Understanding Machine Learning (Shalev-Shwartz and Ben-David)
- Theory of Deep Learning
- Optimization (convex and non-convex)
- And more!

Applied ML

ML applied to basically any problem we might care about (and the tough challenges that come with that)

- Natural Language Processing (NLP)
- Computer Vision
- Computational Biology
- Medical Imaging / Health
- More

Big Picture

Improving the performance at some task through experience!

- Before you start any learning task, remember fundamental questions that will impact how you go about solving it

What is the learning problem?

From what experience?

What model?

What loss function are you optimizing?

With what optimization algorithm?

Are there any guarantees?

How will you evaluate the model?

Who will it impact and how?

Congrats on finishing CSE/STAT 416!
Thanks for the hard work!