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
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!
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

Data

Regression

Intelligence

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

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

Case study: Predicting house prices

Models

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

Including many features:
- Square feet
- # bathrooms
- # bedrooms
- Lot size
- Year built
- ...

\[ \text{Ridge: } w_j \left( \frac{\text{price}}{x} \right) + \lambda \| w \|_2^2 \]
Regression

Case study: Predicting house prices

Algorithms

• Gradient descent

\[ \text{Quality Metric: } \quad \text{RSS}(w) = \sum_{i=1}^{n} (w^T \phi(x_i) - y_i)^2 \]

\[ \hat{w} = \arg \min_w \text{RSS}(w) \]

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

[include all houses]

\[ \hat{w} \]
Regression

Case study: Predicting house prices

Concepts

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

1. Noise
2. Bias
3. Variance

\begin{align*}
\text{Price} & \quad \text{Square feet (sq.ft.)} \\
\hat{y} & \quad X \\
\hat{f}_w & \quad f_w(\text{true}) \\
\end{align*}

Error

Model complexity

Training set
Validation set
Test set

fit \( \hat{\omega} \)

test performance of \( \hat{\omega} \) to select \( \lambda^* \)

assess generalization error of \( \hat{\omega} \)
Case Study 2: Sentiment analysis

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

All reviews:

- **Score(x) < 0**
  - Stars / +
  - Text:
    - **7/21/2015**
    - This is probably my favorite place to eat Japanese in Seattle. My husband and I ordered 4 rolls of sushi, Japanese sandwiches, and the appetizer trio and I wasトラブル the 4 special rolls. I would highly recommend the eel tempura. A wonderful wasabi as recommended by other reviewers was amazing. It's more savory and the wasabi is the perfect amount of flavor for the delicate rolls.

- **Score(x) > 0**
  - Stars / -
  - Text:
    - **6/11/2015**
    - Dining here at the sushi bar made me feel like sitting front row to an amazing performance. We didn't have reservations, bargained down to the 10 after work, got here breathlessly at 5:10pm, and got the last two seats in the place.

- **Score(x) > 0**
  - Stars / -
  - Text:
    - **6/9/2015**
    - I came here having high expectations due to the reviews of this place, but I was totally disappointed. The restaurant is small so do your reservations when you come here. Dishes cost from $8-9.5 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

Score(x) < 0

Score(x) > 0

\( \text{AdaBoost} \)

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

\( \omega_i \leftarrow \text{dataset weights} \)
Classification

Case study: Analyzing sentiment

Algorithms

- Boosting
- Learning from weighted data

- Boosting
- Learning from weighted data

<table>
<thead>
<tr>
<th>Decision Tree</th>
<th>Income&gt;$100K?</th>
<th>Credit history?</th>
<th>Savings&gt;$100K?</th>
<th>Market conditions?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>Bad</td>
<td>Yes</td>
<td>Bad</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>Good</td>
<td>No</td>
<td>Good</td>
</tr>
</tbody>
</table>

- Weighted error:
  - Income>$100K?: 0.2
  - Credit history?: 0.35
  - Savings>$100K?: 0.3
  - Market conditions?: 0.4
Classification

Case study: Analyzing sentiment

Concepts

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

\[ \ell(w_0, w_1, w_2) \]

Accuracy isn't always enough

- Class imbalance
- Unfair treatment

Classifier A
- Best classifier
- Classifier B

STAT/CSE 416: Intro to Machine Learning
Case Study 3: Document retrieval

- Embedding (BoW vs. TF-IDF)
- Distance metrics (Euclidean, Manhattan, Cosine)
Case Study 3+:
Document structuring for retrieval

Data \(\rightarrow\) Clustering \(\rightarrow\) Intelligence

- Data
- Clustering
- Intelligence

Categories:
- SPORTS
- WORLD NEWS
- ENTERTAINMENT
- SCIENCE
Case Study 3++:
Dimensionality reduction

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

Images with thousands or millions of pixels

[STAT/CSE 416: Intro to Machine Learning]
Clustering & Retrieval

Case study: Finding documents

Models

- Nearest neighbors
- Clustering
- Hierarchical clustering

query article

set of nearest neighbors

- Nearest neighbors
- Clustering
- Hierarchical clustering

WORLD NEWS

SPORTS

ENTERTAINMENT

SCIENCE

STAT/CSE 416: Intro to Machine Learning
Clustering & Retrieval

*Case study: Finding documents*

**Algorithms**

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

**Cluster distance**

**Data points**

- Epanechnikov Kernel ($\lambda = 0.2$)

$= ???$

$= ???$

Cluster
distance

- Data points

- House size

- Price ($$)

$= ???$
Clustering & Retrieval

Case study: Finding documents

Concepts

- Distance metrics, kernels, approximation algorithms, dimensionality reduction

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

Rating =

Popularity
co-occurrence matrix

Parameters of model

Featurized MF

\[ X = \text{Rating} \]

\[ X_{ij}^{\text{unknown}} \text{ for white cells} \]

Rows index movies
Columns index users

\[ X = \text{Rating} \]

Parameters of model

\[ L \approx R' \]
Recommender Systems & Matrix Factorization

Case study: Recommending Products

Algorithms

• Coordinate descent

Rating = \approx \begin{bmatrix} L \end{bmatrix} \begin{bmatrix} R' \end{bmatrix}

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

STAT/CSE 416: Intro to Machine Learning
Recommender Systems & Matrix Factorization

Case study: Recommending Products

Concepts

• Matrix completion, cold-start problem

Customers

Products

Customers

Products

X =

Customers

Products

Customers

Products

Xij unknown for white cells

Rows index movies

Columns index users
Case Study 5: Image classification

Data → Deep Learning → Intelligence

Layer 1: 1 → x₁ → Layer 2: 1 → z₁ → y
Layer 2: 1 → x₂ → z₂ → y

Face?
Deep Learning

Case study: Image classification

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

Hyperbolic tangent

ReLU
STAT/CSE 416: Intro to Machine Learning

Training Data → Feature extraction → ML model → Quality metric

x → ŷ → ML algorithm → ŷ

Weight vector ŵ
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 PollEverywhere 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
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:
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!
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
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
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
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?

What model?

With what optimization algorithm?

How will you evaluate the model?

From what experience?

What loss function are you optimizing?

Are there any guarantees?

Who will it impact and how?
Congrats on finishing CSE/STAT 416! Thanks for the hard work!