A PRACTICAL GUIDE TO MACHINE LEARNING
TODAY:

Part 1: ML

Part 2: In-class lab

WEDNESDAY:

Part 1: More on embeddings and classification/regressions

Part 2: Overview of potential quiz questions
MACHINE LEARNING PROBLEMS

(Boosted-) Decision Trees

Supervised Learning

Discrete

classification or categorization

Continuous

regression

Unsupervised Learning

classification

clustering

regression

dimensionality reduction

(Boosted-) Decision Trees

K-Means
Agglomerative clustering
DBScan

PCA
WHAT IS A CLASSIFIER

Apply a prediction function to a feature representation of an image/data-set to get the desired output:

\[
f(\text{apple}) = \text{“apple”}
\]

\[
f(\text{tomato}) = \text{“tomato”}
\]

\[
f(\text{cow}) = \text{“cow”}
\]
y = f(x)

Training: given a training set of labeled examples \( \{(x_1, y_1), \ldots, (x_N, y_N)\} \), estimate the prediction function \( f \) by minimizing the prediction error on the training set.

Testing: apply \( f \) to a never before seen test example \( x \) and output the predicted value \( y = f(x) \).
ML PIPELINE (SUPERVISED)

(a) Training

label

input

feature extractor

features

machine learning algorithm

(b) Prediction

input

feature extractor

features

classifier model

label
MANY CLASSIFIERS TO CHOOSE FROM

K-nearest neighbor
Support Vector Machines
Decision Trees
Random Forrest
(Gradient) Boosted Decision Trees
Logistic Regression
Naïve Bayes
Bayesian network
RBM's

Which is the best one?
MANY CLASSIFIERS TO CHOOSE FROM

K-nearest neighbor
Support Vector Machines
Decision Trees
Random Forrest
(Gradient) Boosted Decision Trees
Logistic Regression
Naïve Bayes
Bayesian network
RBM's

Which is the best one?
CLASSIFIERS: NEAREST NEIGHBOR

\[ f(x) = \text{label of the training example nearest to } x \]

- All we need is a distance function for our inputs
- No training required!
K-NEAREST NEIGHBOR
1-NEAREST NEIGHBOR
3-NEAREST NEIGHBOR
5-NEAREST NEIGHBOR
DECISION BOUNDARIES KNN

Assign label of nearest training data point to each test data point

Voronoi partitioning of feature space for two-category 2D and 3D data

Source: D. Lowe
MANY CLASSIFIERS TO CHOOSE FROM

K-nearest neighbor

Support Vector Machines

Decision Trees

Random Forest

(Gradient) Boosted Decision Trees

Logistic Regression

Naïve Bayes

Bayesian network

RBMs

....

Which is the best one?
Find a _linear function_ to separate the classes:

\[ f(x) = \text{sgn}(w \cdot x + b) \]
LINEAR CLASSIFIERS

\[ f(x, w, b) = \text{sign}(w \cdot x - b) \]

How would you classify this data?
LINEAR CLASSIFIERS

\[ f(x, w, b) = \text{sign}(w \cdot x - b) \]

How would you classify this data?
LINEAR CLASSIFIERS

Denotes +1
Denotes -1

\[ f(x, w, b) = \text{sign}(w \cdot x - b) \]

How would you classify this data?

Slides from Andrew W. Moore
LINEAR CLASSIFIERS

\[ f(x, w, b) = \text{sign}(w \cdot x - b) \]

How would you classify this data?

\( \alpha \)

\( x \rightarrow f \rightarrow y^{\text{est}} \)

- \( \bullet \) denotes +1
- \( \circ \) denotes -1

Slides from Andrew W. Moore
LINEAR CLASSIFIERS

\[ f(x, w, b) = \text{sign}(w \cdot x - b) \]

- \( \text{denotes } +1 \)
- \( \text{denotes } -1 \)

Any of these would be fine..

..but which is best?
Define the **margin** of a linear classifier as the width that the boundary could be increased by before hitting a datapoint.

$$f(x, w, b) = \text{sign}(w \cdot x - b)$$

- denotes +1
- denotes -1

Slides from Andrew W. Moore
The maximum margin linear classifier is the linear classifier with the maximum margin. This is the simplest kind of SVM (Called an LSVM)

\[ f(x, w, b) = \text{sign}(w \cdot x - b) \]
The **maximum margin linear classifier** is the linear classifier with the maximum margin. This is the simplest kind of SVM (Called an LSVM)

\[ f(x, w, b) = \text{sign}(w \cdot x - b) \]

Support Vectors are those datapoints that the margin pushes up against.

\[ f(x, w, b) = \text{sign}(w \cdot x - b) \]

- denotes +1
- denotes -1
The maximum margin linear classifier is the linear classifier with the maximum margin. This is the simplest kind of SVM (Called an LSVM).

\[ f(x, w, b) = \text{sign}(w \cdot x - b) \]

Support vectors are those data points that the margin pushes up against.

- \( +1 \) denotes
- \( -1 \) denotes

sklearn.linear_model.SGDClassifier
Default loss: “hinge” \( \rightarrow \) linear SVM.
MANY CLASSIFIERS TO CHOOSE FROM

K-nearest neighbor
Support Vector Machines
Decision Trees
Random Forrest
(Gradient) Boosted Decision Trees
Logistic Regression
Naïve Bayes
Bayesian network
RBMs
....

Which is the best one?
DECISION TREES

sex

female
satisfy

male
not satisfy

pclass

age

<=4
satisfy

>4
not satisfy

not satisfy

not satisfy
RANDOM FOREST

Sample with Replacement & select random subset of features*

Build classifier over sample

Use majority Vote for classification (or avg. for regression)

* Normally done for each node of the decision tree – not once
Take a set of weak classifiers (normally they should do better than guessing)

Combine to form the Final strong classifier

\[ H(x) = \text{sign} \sum_{i=1}^{n} q_i h_i(x) \]
ADABOOST - CORE IDEA

Take a set of weak classifiers (normally they should do better than guessing)

\[ H(x) = \text{sign} \sum_{i=1}^{n} q_i h_i(x) \]

Classification Result

Weight the result of each classify with
ADABOOST - CORE IDEA

Take a set of weak classifiers (normally they should do better than guessing)

Classification Result

Weight the result of each classify with

\[ H(x) = \text{sign} \sum_{i=1}^{n} q_i h_i(x) \]

Combine to form the Final strong classifier

XGBoost follows the same idea
ML PIPELINE (SUPERVISED)

(a) Training
- Input
- Feature extractor
- Features
- Machine learning algorithm
- Label

(b) Prediction
- Input
- Feature extractor
- Features
- Classifier model
- Label
FEATURES

Fact Table
- Shop_ID
- Customer_ID
- Date_ID
- Product_ID
- Amount
- Volume
- Profit
- ...

Fact Table
- Shop_ID
- Customer_ID
- Date_ID
- Product_ID
- Amount
- Volume
- Profit
- Delivery Time
- ...

Product
- Product_ID
- Type_ID
- Brand_ID
- Length
- Height
- Depth
- Weight
- ...

Product_Type
- Type_ID
- Name
- Description
- ...

Brand
- Brand_ID
- Name
- ...

<table>
<thead>
<tr>
<th>Customer State</th>
<th>Product Type</th>
<th>Product Weight</th>
<th>Volume (L<em>H</em>D)</th>
<th>Month</th>
<th>Delivery Time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Custermer State | Product Type | Product Weight | Volume (L*H*D) | Month | Delivery Time |
IMAGE FEATURES

Raw pixels

Histograms

GIST descriptors

...
FDA approved on-line pharmacies. Chose your product and site below:

Canadian pharmacy - Cialis Soft Tabs - $5.78, Viagra Professional - $1.38, Human Growth Hormone - $43.37, Meridia - $3.32, Tramadol

HerbalKing - Herbal pills for hair enlargement. Techniques, pro dangerous pumps, exercises and surgeries.

Anatrim - Are you ready for Summer? Use Anatrim, the most pow...
ONE-HOT ENCODING

Bag of Words

\[
\begin{pmatrix}
Viagra \\
Soft \\
Herbel \\
Pills \\
Are \\
\vdots
\end{pmatrix}
\]

<table>
<thead>
<tr>
<th>ID</th>
<th>Viagra</th>
<th>Soft</th>
<th>Herbel</th>
<th>Pills</th>
<th>Are</th>
<th>....</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mail1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>....</td>
</tr>
<tr>
<td>Mail2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>....</td>
</tr>
<tr>
<td>....</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>....</td>
</tr>
<tr>
<td>....</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>....</td>
</tr>
<tr>
<td>....</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>....</td>
</tr>
<tr>
<td>....</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>....</td>
</tr>
<tr>
<td>....</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>....</td>
</tr>
</tbody>
</table>
HOW WOULD YOU ENCODE THE TABLE?

<table>
<thead>
<tr>
<th>Name</th>
<th>ZipCode</th>
<th>Age</th>
<th>Sex</th>
<th>Area</th>
<th>Avg Grade</th>
<th>Statement</th>
<th>Early admit</th>
<th>Accepted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mike</td>
<td>02474</td>
<td>23</td>
<td>M</td>
<td>DB</td>
<td>B-</td>
<td>Since I was born, I knew I wanted to code. My first program I wrote in binary code literally in the sandbox, though I am not sure it was correct...</td>
<td>No</td>
<td>NO</td>
</tr>
<tr>
<td>Sam</td>
<td>02456</td>
<td>21</td>
<td>M</td>
<td>Sens or</td>
<td>A</td>
<td>Celine Dion’s song “A New Day Has Come” taught me that CS is the best subject in the world. I never felt...</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Amadou</td>
<td>15106</td>
<td>22</td>
<td>M</td>
<td>DB</td>
<td>A+</td>
<td>I want to get out of Pittsburgh.</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Anna</td>
<td>02319</td>
<td>22</td>
<td>F</td>
<td>ML</td>
<td>A-</td>
<td>I already wrote 10 papers and I think I am ready to graduate now.</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Name</td>
<td>ZipCode</td>
<td>Age</td>
<td>Sex</td>
<td>Area</td>
<td>Avg Grade</td>
<td>Statement</td>
<td>Early admit</td>
<td>Accepted</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>-----</td>
<td>-----</td>
<td>------</td>
<td>-----------</td>
<td>---------------------------------------------------------------------------</td>
<td>-------------</td>
<td>----------</td>
</tr>
<tr>
<td>Mike</td>
<td>02474</td>
<td>23</td>
<td>M</td>
<td>DB</td>
<td>B-</td>
<td>Since I was born, I knew I wanted to code. My first program I wrote in</td>
<td>No</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>binary code literally in the sandbox, though I am not sure it was correct...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sam</td>
<td>02456</td>
<td>21</td>
<td>null</td>
<td>Sens or</td>
<td>A</td>
<td>Celine Dion’s song “A New Day Has Come” taught me that CS is the best</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>subject in the world. I never felt...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amadou</td>
<td>15106</td>
<td>22</td>
<td>M</td>
<td>DB</td>
<td>A+</td>
<td>I want to get out of Pittsburgh.</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Anna</td>
<td>null</td>
<td>22</td>
<td>F</td>
<td>ML</td>
<td>A-</td>
<td>I already wrote 10 papers and I think I am ready to graduate now.</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Remove identifiers
Encode as (1) Lat/Lon and scale to 0-1, or remove
Scale to 0-1
1-Hot Encode or remove

Bag of words
1-Hot Encoding
Remove (information leakage)
## PREDICTOR FOR GRAD-SCHOOL APPLICATIONS

<table>
<thead>
<tr>
<th>Name</th>
<th>ZipCode</th>
<th>Age</th>
<th>Sex</th>
<th>Area</th>
<th>Avg Grade</th>
<th>Statement</th>
<th>Early admit</th>
<th>Accepted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mike</td>
<td>02474</td>
<td>23</td>
<td>M</td>
<td>DB</td>
<td>B-</td>
<td>Since I was born, I knew I wanted to code. My first program I wrote in binary code literally in the sandbox, though I am not sure it was correct...</td>
<td>No</td>
<td>NO</td>
</tr>
<tr>
<td>Sam</td>
<td>02456</td>
<td>21</td>
<td>null</td>
<td>Sens or</td>
<td>A</td>
<td>Celine Dion’s song “A New Day Has Come” taught me that CS is the best subject in the world. I never felt...</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Amadou</td>
<td>15106</td>
<td>22</td>
<td>M</td>
<td>DB</td>
<td>A+</td>
<td>I want to get out of Pittsburgh.</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Anna</td>
<td>null</td>
<td>22</td>
<td>F</td>
<td>ML</td>
<td>A-</td>
<td>I already wrote 10 papers and I think I am ready to graduate now.</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No zip</th>
<th>Lon</th>
<th>Lat</th>
<th>Age</th>
<th>M</th>
<th>F</th>
<th>Other</th>
<th>ML</th>
<th>DB</th>
<th>Sens</th>
<th>Avg Grade</th>
<th>Born</th>
<th>Celine</th>
<th>Pittsburg</th>
<th>Accepted</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.123</td>
<td>.0222</td>
<td>0.8</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.5</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NO</td>
</tr>
<tr>
<td>0</td>
<td>.123</td>
<td>.0123</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>M</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>Yes</td>
</tr>
<tr>
<td>0</td>
<td>.031</td>
<td>.0322</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>M</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>