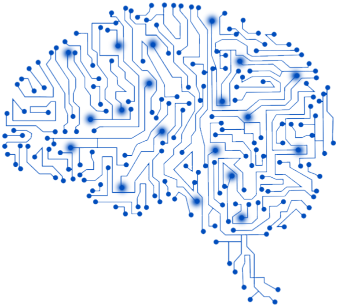
Machine Learning 2018 Final Project

Taha Azzaoui, Daniel Santos, Huma Sheikh, Sovanmony Lim

University of Massachusetts, Lowell

December 10, 2018



Labeled Faces in the Wild: Overview

The Data set

- 13233 images, 5749 people
- 150px $\times$ 150px

Classification Problem

- $Y = \{$George Bush, Colin Powell, ...$\}$ (labels)
- $X = \{(x, y) | x \in \mathbb{R}^{150\times150}, y \in Y\}$ (labeled images)
- Given an input image $(x \notin X) \rightarrow$ the name of the person $(y \in Y)$

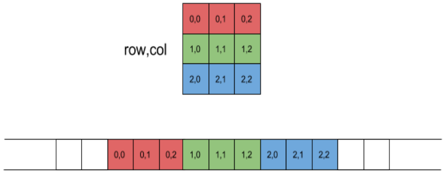
Labeled Faces in the Wild: Initial Attempt

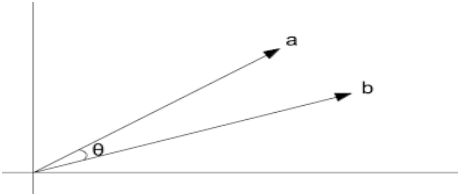
Naive Approach

- Define $f : \mathbb{R}^{150\times150} \to \mathbb{R}^{22500\times1}$
- "Flatten" out images and compute pairwise cosine similarity across entire training set
- $sim(\vec{u}, \vec{v}) = cos(\theta) = \frac{\vec{u}\cdot\vec{v}}{||\vec{v}||||\vec{u}||}$

Problems

- Scale: Computing pairwise cosine similarity across entire database is prohibitively expensive at scale
- Noise: Images are noisy $\implies$ not all pixels matter





Labeled Faces in the Wild: Refined Attempt

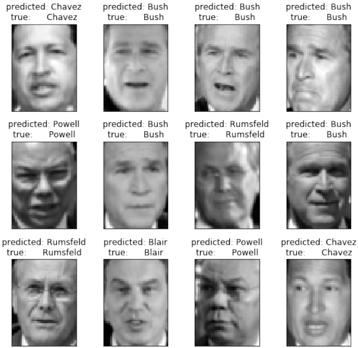
Dimensionality Reduction To the Rescue!

- $g : \mathbb{R}^{22500\times 1} \rightarrow \mathbb{R}^{k\times 1}; k << 22500$
- Idea: Since not all pixels matter, use PCA to find the ones that do.
- Solves the problem of scale and more robust to noise

Support Vector Machines

- Feed $g \circ f$ into an SVM classifier
- Nonlinear decision boundary (deg $= 3$)

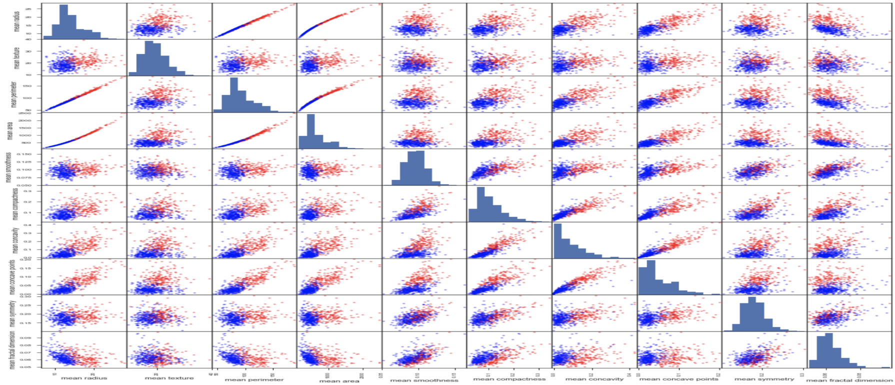
	precision	recall	f1-score	support
Ariel Sharon	0.90	0.75	0.82	24
Colin Powell	0.80	0.91	0.85	58
Donald Rumsfeld	0.87	0.76	0.81	34
George W Bush	0.83	0.93	0.88	129
Gerhard Schroeder	0.86	0.75	0.80	24
Hugo Chavez	0.92	0.67	0.77	18
Tony Blair	0.93	0.74	0.83	35
micro avg	0.85	0.85	0.85	322
macro avg	0.87	0.79	0.82	322
weighted avg	0.85	0.85	0.85	322



Classification: Breast Cancer

The data set

- 30 features and 569 instances
- Contain radius, textures, perimeter, area, and more
- Classes: Malignant, Benign



Classification: Breast Cancer

Solutions

- Decision Tree
  - Leaves: Classes
  - Branches: Logical Conjunction of Features
  - Average: 0.93
- Logistic Regression
  - Average: 0.98
- Support Vector Machine
  - $\gamma = 0.00001$
  - $C = 10000$
  - Average: 0.98

Can we do better?

Classification: Breast Cancer

Multi-Layer Perceptron Classifier (Neural Network)

- Design
  - 3 Hidden Layers: (20, 10, 30)
  - Activation: Rectified Linear Unit (ReLU)
  - Optimizer: Adam Optimization (Extension to SGD)

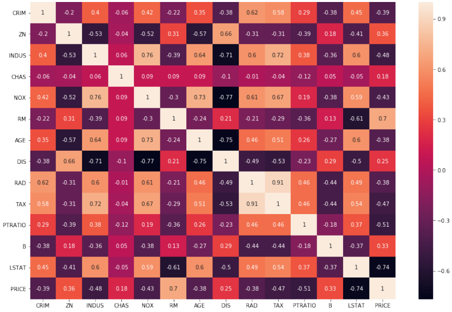
Table: MLP Classification Report

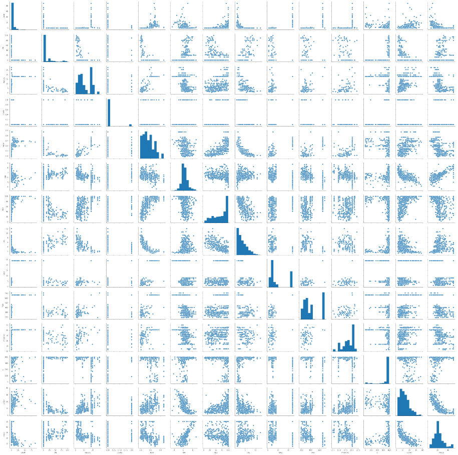
	precision	recall	f1-score	support
WDBC-Malignant	1.00	0.97	0.98	61
WDBC-Benign	0.98	1.00	0.99	110
avg / total	0.98	0.99	0.99	171

- Better, but is it worth?

Regression: Boston Housing Data Set

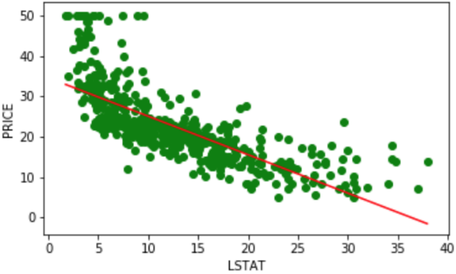
- 506 Samples, 13 features
- Idea: Compare and contrast Neural Network and Linear Regression
- Question: Which features are most correlated with housing prices?

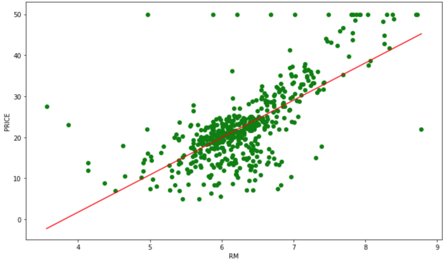




Linear Regression

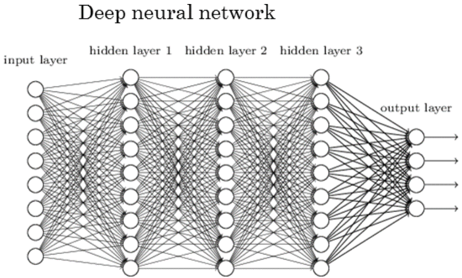
- Number of Rooms & percentage lower status of the population well correlated with housing prices
- Advantages
  - Simple model
  - Easy to train
  - Works well for linear relationships in data
- Mean Squared Error = 21.89.





Neural Network

- How do more sophisticated models compare?
- Architecture: 3 hidden layers
  - Layer 1 = 50 Nodes, Layer 2 = 100 Nodes, Layer 3 = 50 Nodes
  - Activation Function: RELU + Adam Optimizer
  - 50 Epochs
- Mean Squared Error = 19.3



Questions?