Lecture 2: Image Classification pipeline
Administrative

First assignment will come out tonight (or tomorrow at worst)
It is due **January 20** (i.e. in two weeks). Handed in through CourseWork
It includes:
- Write/train/evaluate a kNN classifier
- Write/train/evaluate a Linear Classifier (SVM and Softmax)
- Write/train/evaluate a 2-layer Neural Network (backpropagation!)
- Requires writing numpy/Python code

**Warning**: don’t work on assignments from last year!

Compute: Can use your own laptops, or Terminal.com
distances = np.sum(np.abs(self.Xtr - X[i,:]), axis = 1)
min_index = np.argmin(distances) # get the index with smallest distance
Ypred[i] = self.ytr[min_index] # predict the label of the nearest example
Image Classification: a core task in Computer Vision

(assume given set of discrete labels)
\{dog, cat, truck, plane, ...\}

cat
The problem: semantic gap

Images are represented as 3D arrays of numbers, with integers between [0, 255].

E.g.
300 x 100 x 3

(3 for 3 color channels RGB)
Challenges: Viewpoint Variation
Challenges: Illumination
Challenges: Deformation
Challenges: Occlusion
Challenges: Background clutter
Challenges: Intraclass variation
An image classifier

Unlike e.g. sorting a list of numbers,

**no obvious way** to hard-code the algorithm for recognizing a cat, or other classes.
Attempts have been made
Data-driven approach:
1. Collect a dataset of images and labels
2. Use Machine Learning to train an image classifier
3. Evaluate the classifier on a withheld set of test images

```python
def train(train_images, train_labels):
    # build a model for images -> labels...
    return model

def predict(model, test_images):
    # predict test_labels using the model...
    return test_labels
```
First classifier: **Nearest Neighbor Classifier**

```python
def train(train_images, train_labels):
    # build a model for images -> labels...
    return model

def predict(model, test_images):
    # predict test_labels using the model...
    return test_labels
```

Remember all training images and their labels

Predict the label of the most similar training image
Example dataset: **CIFAR-10**

- **10 labels**
- **50,000** training images, each image is tiny: 32x32
- **10,000** test images.

![CIFAR-10 Images](image_url)
Example dataset: CIFAR-10
10 labels
50,000 training images
10,000 test images.

For every test image (first column), examples of nearest neighbors in rows
How do we compare the images? What is the distance metric?

**L1 distance:**

\[ d_1(I_1, I_2) = \sum_p |I_1^p - I_2^p| \]

<table>
<thead>
<tr>
<th>test image</th>
<th>training image</th>
<th>pixel-wise absolute value differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>56 32 10 18</td>
<td>10 20 24 17</td>
<td>46 12 14 1</td>
</tr>
<tr>
<td>90 23 128 133</td>
<td>8 10 89 100</td>
<td>82 13 39 33</td>
</tr>
<tr>
<td>24 26 178 200</td>
<td>12 16 178 170</td>
<td>12 10 0 30</td>
</tr>
<tr>
<td>2 0 255 220</td>
<td>4 32 233 112</td>
<td>2 32 22 108</td>
</tr>
</tbody>
</table>

\[ \text{add} \rightarrow 456 \]
import numpy as np

class NearestNeighbor:
    def __init__(self):
        pass

    def train(self, X, y):
        """ X is N x D where each row is an example. Y is 1-dimension of size N ""
        # the nearest neighbor classifier simply remembers all the training data
        self.Xtr = X
        self.ytr = y

    def predict(self, X):
        """ X is N x D where each row is an example we wish to predict label for ""
        num_test = X.shape[0]
        # let's make sure that the output type matches the input type
        Ypred = np.zeros(num_test, dtype = self.ytr.dtype)

        # loop over all test rows
        for i in xrange(num_test):
            # find the nearest training image to the i'th test image
            # using the L1 distance (sum of absolute value differences)
            distances = np.sum(np.abs(self.Xtr - X[i,:]), axis = 1)
            min_index = np.argmin(distances) # get the index with smallest distance
            Ypred[i] = self.ytr[min_index] # predict the label of the nearest example

        return Ypred
Nearest Neighbor classifier

```python
class NearestNeighbor:
    pass

def train(self, X, y):
    # X is NxD where each row is an example. Y is 1-dimensional of size N
    train = X, y
    return train

def predict(self, X):
    # lets make sure that the output type matches the input type
    ypred = np.zeros((num_test, dtype = self.y.dtype))
    for i in range(num_test):
        # find the nearest training image to the i-th test image
        # using the L1 distance (sum of absolute value differences)
        distances = np.sum(np.abs(self.Xt - X[i,:]), axis = 1)
        min_index = np.argmin(distances)
        # get the index with smallest distance
        ypred[i] = self.Yt[min_index]
    # return the index with smallest distance
    return ypred
```

remember the training data
Nearest Neighbor classifier

for every test image:
- find nearest train image with L1 distance
- predict the label of nearest training image

```python
import numpy as np

class NearestNeighbor:
    def __init__(self):
        pass

    def train(self, X, y):
        """X is N x D where each row is an example. Y is 1-dimension of size N""
        # the nearest neighbor classifier simply remembers all the training data
        self.Xtr = X
        self.ytr = y

    def predict(self, X):
        """X is N x D where each row is an example we wish to predict label for""
        num_test = X.shape[0]
        # lets make sure that the output type matches the input type
        Ypred = np.zeros(num_test, dtype = self.ytr.dtype)

        # loop over all test rows
        for i in xrange(num_test):
            # find the nearest training image to the i'th test image
            # using the L1 distance (sum of absolute value differences)
            distances = np.sum(np.abs(self.Xtr - X[i,:]), axis = 1)
            min_index = np.argmin(distances) # get the index with smallest distance
            Ypred[i] = self.ytr[min_index] # predict the label of the nearest example

        return Ypred
```
Nearest Neighbor classifier

Q: how does the classification speed depend on the size of the training data?

```python
class NearestNeighbor:
    def __init__(self):
        pass

def train(self, X, y):
    """ X is N x D where each row is an example. Y is 1-dimensional of size N ""
    # the nearest neighbor classifier simply remembers all the training data
    self.Xtr = X
    self.ytr = y

def predict(self, X):
    """ X is N x D where each row is an example we wish to predict label for ""
    num_test = X.shape[0]
    # lets make sure that the output type matches the input type
    Ypred = np.zeros(num_test, dtype = self.ytr.dtype)

    # loop over all test rows
    for i in xrange(num_test):
        # find the nearest training image to the i'th test image
        # using the L1 distance (sum of absolute value differences)
        distances = np.sum(np.abs(self.Xtr - X[i,:]), axis = 1)
        min_index = np.argmin(distances) # get the index with smallest distance
        Ypred[i] = self.ytr[min_index] # predict the label of the nearest example

    return Ypred
```
import numpy as np

class NearestNeighbor:
    def __init__(self):
        pass

    def train(self, X, y):
        # X is N x D where each row is an example. Y is 1-dimension of size N
        # the nearest neighbor classifier simply remembers all the training data
        self.Xtr = X
        self.ytr = y

    def predict(self, X):
        # X is N x D where each row is an example we wish to predict label for
        num_test = X.shape[0]
        Ypred = np.zeros(num_test, dtype = self.ytr.dtype)

        # loop over all test rows
        for i in xrange(num_test):
            # find the nearest training image to the i'th test image
            # using the L1 distance (sum of absolute value differences)
            distances = np.sum(np.abs(self.Xtr - X[i,:]), axis = 1)
            min_index = np.argmin(distances)  # get the index with smallest distance
            Ypred[i] = self.ytr[min_index]  # predict the label of the nearest example

        return Ypred

Nearest Neighbor classifier

Q: how does the classification speed depend on the size of the training data? linearly :

This is backwards:
- test time performance is usually much more important in practice.
- CNNs flip this: expensive training, cheap test evaluation
Aside: Approximate Nearest Neighbor
find approximate nearest neighbors quickly

ANN: A Library for Approximate Nearest Neighbor Searching
David M. Mount and Sunil Arya
Version 1.1.2
Release Date: Jan 27, 2010

What is ANN?
ANN is a library written in C++, which supports data structures and algorithms for both exact and approximate nearest neighbor searching in arbitrary high dimensions.

In the nearest neighbor problem a set of data points in d-dimensional space is given. These points are preprocessed into a data structure, so that given any query point q, the nearest or generally k nearest points of P to q can be reported efficiently. The distance between two points can be defined in many ways. ANN assumes that distances are measured using any class of distance functions called Minkowski metrics. These include the well known Euclidean distance, Manhattan distance, and max distance.

Based on our own experience, ANN performs quite efficiently for point sets ranging in size from thousands to hundreds of thousands, and in dimensions as high as 20. (For applications in significantly higher dimensions, the results are rather spotty, but you might try it anyway.)

The library implements a number of different data structures, based on kd-trees and box-decomposition trees, and employs a couple of different search strategies.

The library also comes with test programs for measuring the quality of performance of ANN on any particular data sets, as well as programs for visualizing the structure of the geometric data structures.
The choice of distance is a hyperparameter common choices:

L1 (Manhattan) distance

\[ d_1(I_1, I_2) = \sum_p |I^P_1 - I^P_2| \]

L2 (Euclidean) distance

\[ d_2(I_1, I_2) = \sqrt{\sum_p (I^P_1 - I^P_2)^2} \]
k-Nearest Neighbor
find the k nearest images, have them vote on the label

Example dataset: CIFAR-10
10 labels
50,000 training images
10,000 test images.

For every test image (first column), examples of nearest neighbors in rows.
Q: what is the accuracy of the nearest neighbor classifier on the training data, when using the Euclidean distance?
Q2: what is the accuracy of the $k$-nearest neighbor classifier on the training data?
What is the best **distance** to use? What is the best value of **k** to use?

i.e. how do we set the **hyperparameters**?
What is the best \textit{distance} to use?
What is the best value of \textit{k} to use?

i.e. how do we set the \textit{hyperparameters}?

Very problem-dependent.
Must try them all out and see what works best.
Try out what hyperparameters work best on test set.
Trying out what hyperparameters work best on test set:
Very bad idea. The test set is a proxy for the generalization performance!
Use only **VERY SPARINGLY**, at the end.
Validation data
use to tune hyperparameters
Cross-validation cycle through the choice of which fold is the validation fold, average results.
Example of 5-fold cross-validation for the value of $k$.

Each point: single outcome.

The line goes through the mean, bars indicated standard deviation

(Seems that $k \approx 7$ works best for this data)
k-Nearest Neighbor on images never used.

- terrible performance at test time
- distance metrics on level of whole images can be very unintuitive

(all 3 images have same L2 distance to the one on the left)
Summary

- **Image Classification**: We are given a **Training Set** of labeled images, asked to predict labels on **Test Set**. Common to report the **Accuracy** of predictions (fraction of correctly predicted images).
- We introduced the **k-Nearest Neighbor Classifier**, which predicts the labels based on nearest images in the training set.
- We saw that the choice of distance and the value of k are **hyperparameters** that are tuned using a **validation set**, or through **cross-validation** if the size of the data is small.
- Once the best set of hyperparameters is chosen, the classifier is evaluated once on the test set, and reported as the performance of kNN on that data.
Linear Classification
Neural Networks practitioner
"man in black shirt is playing guitar."
"construction worker in orange safety vest is working on road."
"two young girls are playing with lego toy."
"boy is doing backflip on wakeboard."
"girl in pink dress is jumping in air."
"black and white dog jumps over bar."
"young girl in pink shirt is swinging on swing."
"man in blue wetsuit is surfing on wave."
CNN

RNN

“straw”  “hat”  END

\[ y_t \]

\[ W_{hh} \]

\[ W_{oh} \]

\[ h_t \]

\[ x_t \]

\[ W_{hx} \]
Example dataset: CIFAR-10
10 labels
50,000 training images
each image is 32x32x3
10,000 test images.
Parametric approach

$[32\times32\times3]$ array of numbers $0...1$ (3072 numbers total)

$f(x, W)$

image parameters

10 numbers, indicating class scores
Parametric approach: **Linear classifier**

\[ f(x, W) = Wx \]

[32x32x3] array of numbers 0...1

10 numbers, indicating class scores
Parametric approach: Linear classifier

\[ f(x, W) = Wx \]

[\[32x32x3\]]
array of numbers 0...1

10x1 10x3072

3072x1

10 numbers, indicating class scores

parameters, or “weights”
Parametric approach: **Linear classifier**

![Diagram](image)

\[ f(x, W) = Wx \]

- **[32x32x3]** array of numbers 0...1
- 10x1 parameters, or “weights”
- 10x3072
- 3072x1
- (+b) 10x1
- 10 numbers, indicating class scores
Example with an image with 4 pixels, and 3 classes (cat/dog/ship)
Interpreting a Linear Classifier

Q: what does the linear classifier do, in English?

\[ f(x_i, W, b) = Wx_i + b \]
Interpreting a Linear Classifier

Example trained weights of a linear classifier trained on CIFAR-10:
Interpreting a Linear Classifier

$$f(x_i, W, b) = WX_i + b$$

[32x32x3] array of numbers 0...1 (3072 numbers total)
Interpreting a Linear Classifier

Q2: what would be a very hard set of classes for a linear classifier to distinguish?

\[ f(x_i, W, b) = W x_i + b \]
So far: We defined a (linear) score function: \( f(x_i, W, b) = Wx_i + b \)

Example class scores for 3 images, with a random \( W \):

<table>
<thead>
<tr>
<th></th>
<th>score</th>
<th>score</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>airplane</td>
<td>-3.45</td>
<td>-0.51</td>
<td>3.42</td>
</tr>
<tr>
<td>automobile</td>
<td>-8.87</td>
<td>6.04</td>
<td>4.64</td>
</tr>
<tr>
<td>bird</td>
<td>0.09</td>
<td>5.31</td>
<td>2.65</td>
</tr>
<tr>
<td>cat</td>
<td>2.9</td>
<td>-4.22</td>
<td>5.1</td>
</tr>
<tr>
<td>deer</td>
<td>4.48</td>
<td>-4.19</td>
<td>2.64</td>
</tr>
<tr>
<td>dog</td>
<td>8.02</td>
<td>3.58</td>
<td>5.55</td>
</tr>
<tr>
<td>frog</td>
<td>3.78</td>
<td>4.49</td>
<td>-4.34</td>
</tr>
<tr>
<td>horse</td>
<td>1.06</td>
<td>-4.37</td>
<td>-1.5</td>
</tr>
<tr>
<td>ship</td>
<td>-0.36</td>
<td>-2.09</td>
<td>-4.79</td>
</tr>
<tr>
<td>truck</td>
<td>-0.72</td>
<td>-2.93</td>
<td>6.14</td>
</tr>
</tbody>
</table>
Coming up:

- Loss function
  (quantifying what it means to have a “good” $W$)
- Optimization
  (start with random $W$ and find a $W$ that minimizes the loss)
- ConvNets!
  (tweak the functional form of $f$)