Assignment 3

Unsupervised Learning and Probabilistic Models

Jimmy Ba
jimmy@psi.toronto.edu

Background

- Unsupervised learning plays central role in a wide range of computer science and engineering problems

"People read around 10 MB worth of material a day, hear 400 MB a day, and see 1 MB of information every second" - The Economist, November 2006

In 2015, consumption will raise to 74 GB a day - UCSD Study 2014

Motivation

- Getting meaningful representations from text/image/sensor data is often the key component in Google search engine or your next big start-up ideas

Motivation

- Getting meaningful representations from text/image/sensor data is often the key component in Google search engine or your next big start-up ideas

- The most common unsupervised learning problem: Recommendations

Clustering

- Recommendation problem is a particular instance of clustering

- The fundamental idea is to assume not all data points are created equal. Some data points are more similar than others.

- We would like to discover the “prototypes” or cluster centres that summarize the underlying dataset

Clustering

- Some 2D data scatter plots

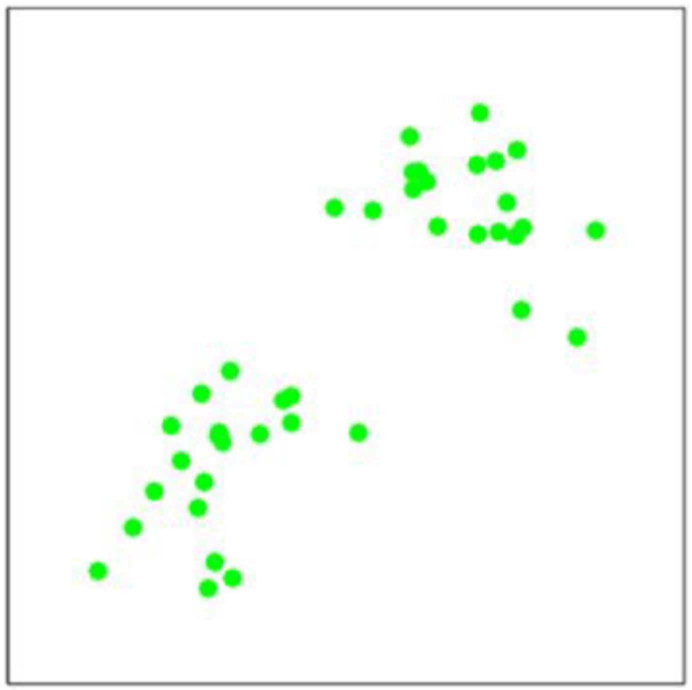


figure credit: Andrew Ng, cs229-notes7a, 2013

Clustering

- Some 2D data scatter plots
- Looks like there are two clusters

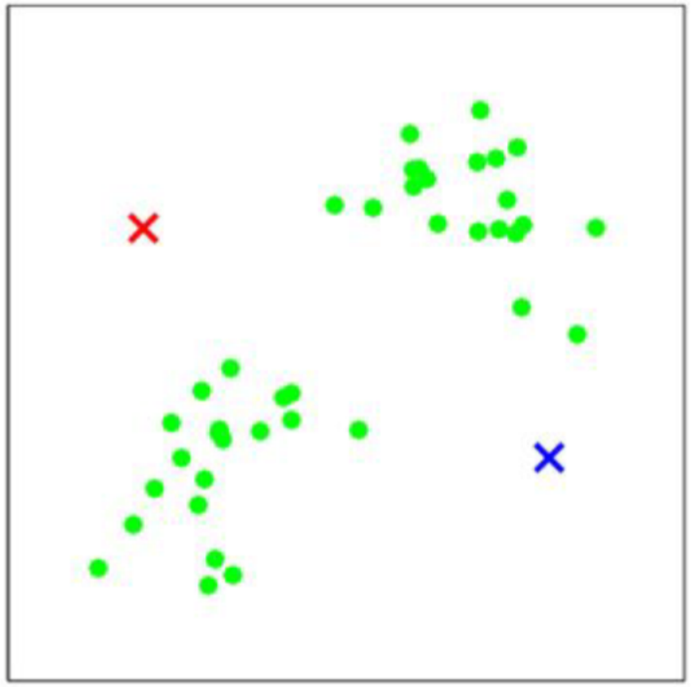


figure credit: Andrew Ng, cs229-notes7a, 2013

Clustering

- Some 2D data scatter plots
- Looks like there are two clusters
- Assign data points to the current cluster centres

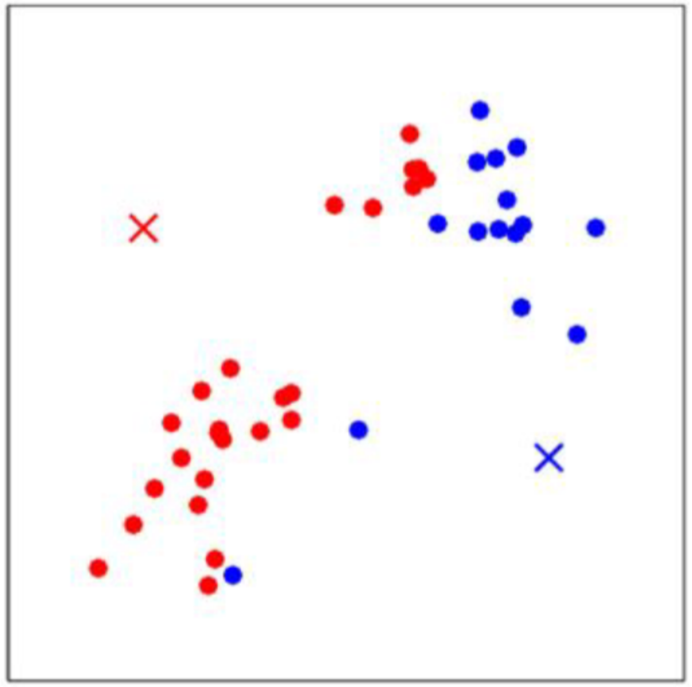


figure credit: Andrew Ng, cs229-notes7a, 2013

Clustering

- Some 2D data scatter plots
- Looks like there are two clusters
- Assign data points to the current cluster centres
- Update cluster

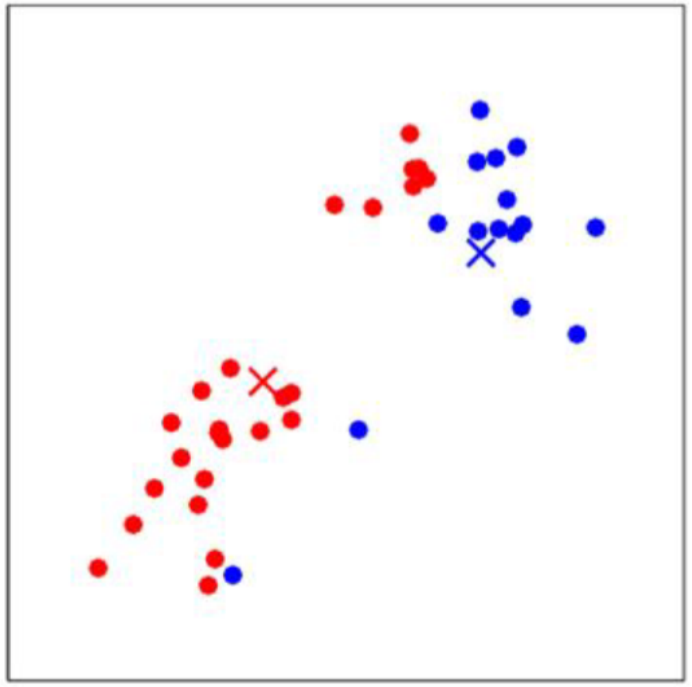


figure credit: Andrew Ng, cs229-notes7a, 2013

Clustering

- Some 2D data scatter plots

- Looks like there are two clusters

- Assign data points to the current cluster centres

- Update cluster

repeat

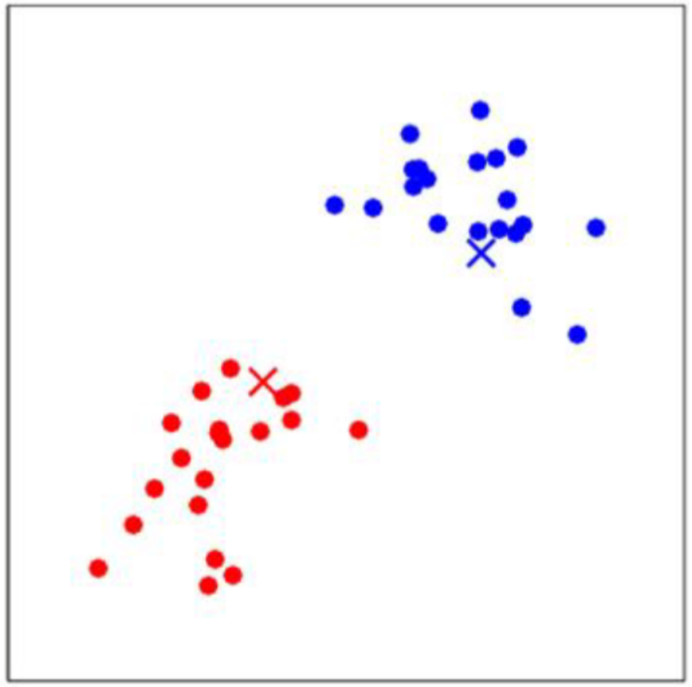


figure credit: Andrew Ng, cs229-notes7a, 2013

Clustering

- Some 2D data scatter plots
- Looks like there are two clusters
- Assign data points to the current cluster centres
- Update cluster

repeat

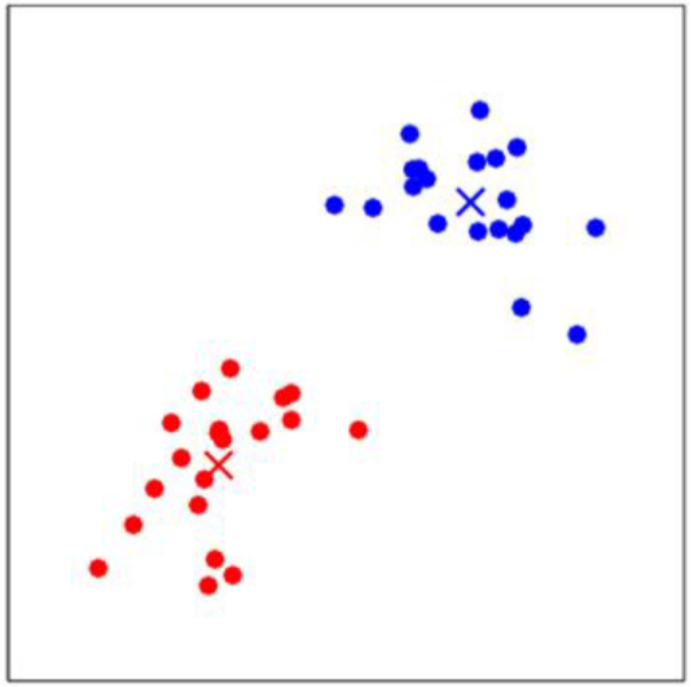


figure credit: Andrew Ng, cs229-notes7a, 2013

K-means clustering algorithm

1. Initialize **cluster centroids** $\mu_1, \mu_2, \ldots, \mu_k \in \mathbb{R}^n$ randomly.

2. Repeat until convergence: {

   For every i, set

$$c^{(i)} := \arg\min_j ||x^{(i)} - \mu_j||^2.$$

   For each j, set

$$\mu_j := \frac{\sum_{i=1}^m 1\{c^{(i)} = j\} x^{(i)}}{\sum_{i=1}^m 1\{c^{(i)} = j\}}.$$

}

figure credit: Andrew Ng, cs229-notes7a, 2013

K-means clustering algorithm

1. Initialize **cluster centroids** $\mu_1, \mu_2, \ldots, \mu_k \in \mathbb{R}^n$ randomly.

2. Repeat until convergence: {

   For every i, set

   Cluster assignment step

   $$c^{(i)} := \arg\min_j ||x^{(i)} - \mu_j||^2.$$

   For each j, set

   $$\mu_j := \frac{\sum_{i=1}^m 1\{c^{(i)} = j\} x^{(i)}}{\sum_{i=1}^m 1\{c^{(i)} = j\}}.$$

}

figure credit: Andrew Ng, cs229-notes7a, 2013

K-means clustering algorithm

1. Initialize **cluster centroids** $\mu_1, \mu_2, \ldots, \mu_k \in \mathbb{R}^n$ randomly.

2. Repeat until convergence: $\{$

For every i, set

Cluster assignment step

$$c^{(i)} := \arg\min_j ||x^{(i)} - \mu_j||^2.$$

For each j, set

Cluster updates step

$$\mu_j := \frac{\sum_{i=1}^m 1\{c^{(i)} = j\} x^{(i)}}{\sum_{i=1}^m 1\{c^{(i)} = j\}}.$$

$\}$

figure credit: Andrew Ng, cs229-notes7a, 2013

K-means clustering algorithm

1. Initialize **cluster centroids** $\mu_1, \mu_2, \ldots, \mu_k \in \mathbb{R}^n$ randomly.

2. Repeat until convergence: {

For every i, set

Cluster assignment step

$$c^{(i)} := \arg\min_j ||x^{(i)} - \mu_j||^2.$$

For each j, set

Cluster updates step

$$\mu_j := \frac{\sum_{i=1}^m 1\{c^{(i)} = j\} x^{(i)}}{\sum_{i=1}^m 1\{c^{(i)} = j\}}.$$

Homework question:
Why does this algorithm terminate?

}

figure credit: Andrew Ng, cs229-notes7a, 2013

K-means clustering algorithm

- K-means loss function

$$\mathcal{L}(\boldsymbol{\mu}) = \sum_{n=1}^{B} \min_{k=1}^{K} \|\mathbf{x}_n - \boldsymbol{\mu}_k\|_2^2$$

Things you need to do in assignment 1

- K-means loss function

Find this guy

$$\mathcal{L}(\boldsymbol{\mu}) = \sum_{n=1}^{B} \min_{k=1}^{K} \|\mathbf{x}_n - \boldsymbol{\mu}_k\|_2^2$$

Things you need to do in assignment 1

- K-means loss function
- Part 1
  - compute distances

$$\mathcal{L}(\boldsymbol{\mu}) = \sum_{n=1}^{B} \min_{k=1}^{K} \|\mathbf{x}_n - \boldsymbol{\mu}_k\|_2^2$$

Compute squared Euclidean distance

Things you need to do in assignment 1

- K-means loss function
- Part 1
  - compute distances

$$\mathcal{L}(\boldsymbol{\mu}) = \sum_{n=1}^{B} \min_{k=1}^{K} \|\mathbf{x}_n - \boldsymbol{\mu}_k\|_2^2$$

Compute squared Euclidean distance

- ((x-mu)**2).sum()   works for vectors

Things you need to do in assignment 3

- K-means loss function

- Part 1

  - compute distances

$$\mathcal{L}(\boldsymbol{\mu}) = \sum_{n=1}^{B} \min_{k=1}^{K} \|\mathbf{x}_n - \boldsymbol{\mu}_k\|_2^2$$

Compute squared Euclidean distance

- ((x-mu)**2).sum()   works for vectors

How about matrices?

Things you need to do in assignment 3

- K-means loss function
- Part 1
  - compute distances

$$\mathcal{L}(\boldsymbol{\mu}) = \sum_{n=1}^{B} \min_{k=1}^{K} \|\mathbf{x}_n - \boldsymbol{\mu}_k\|_2^2$$

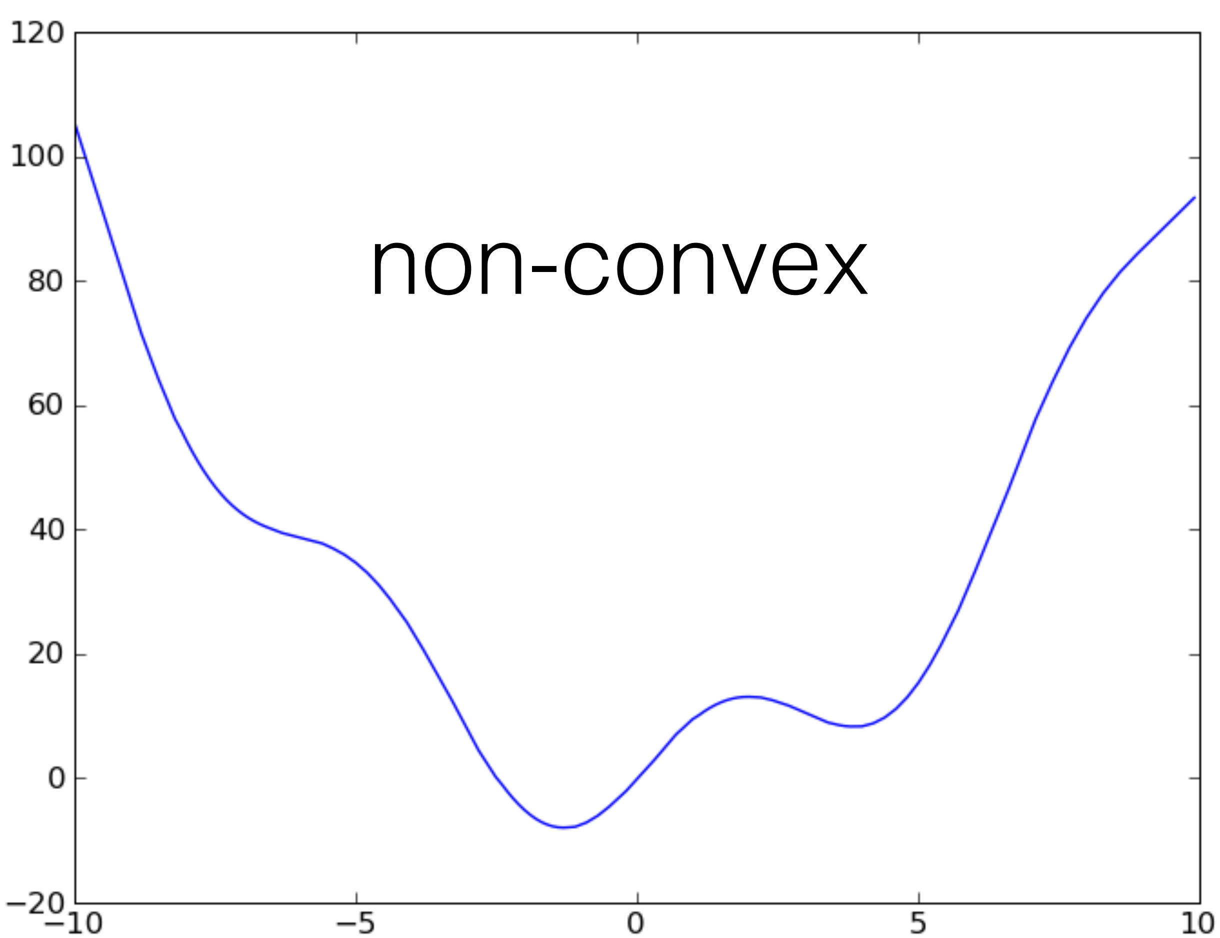
Compute squared Euclidean distance

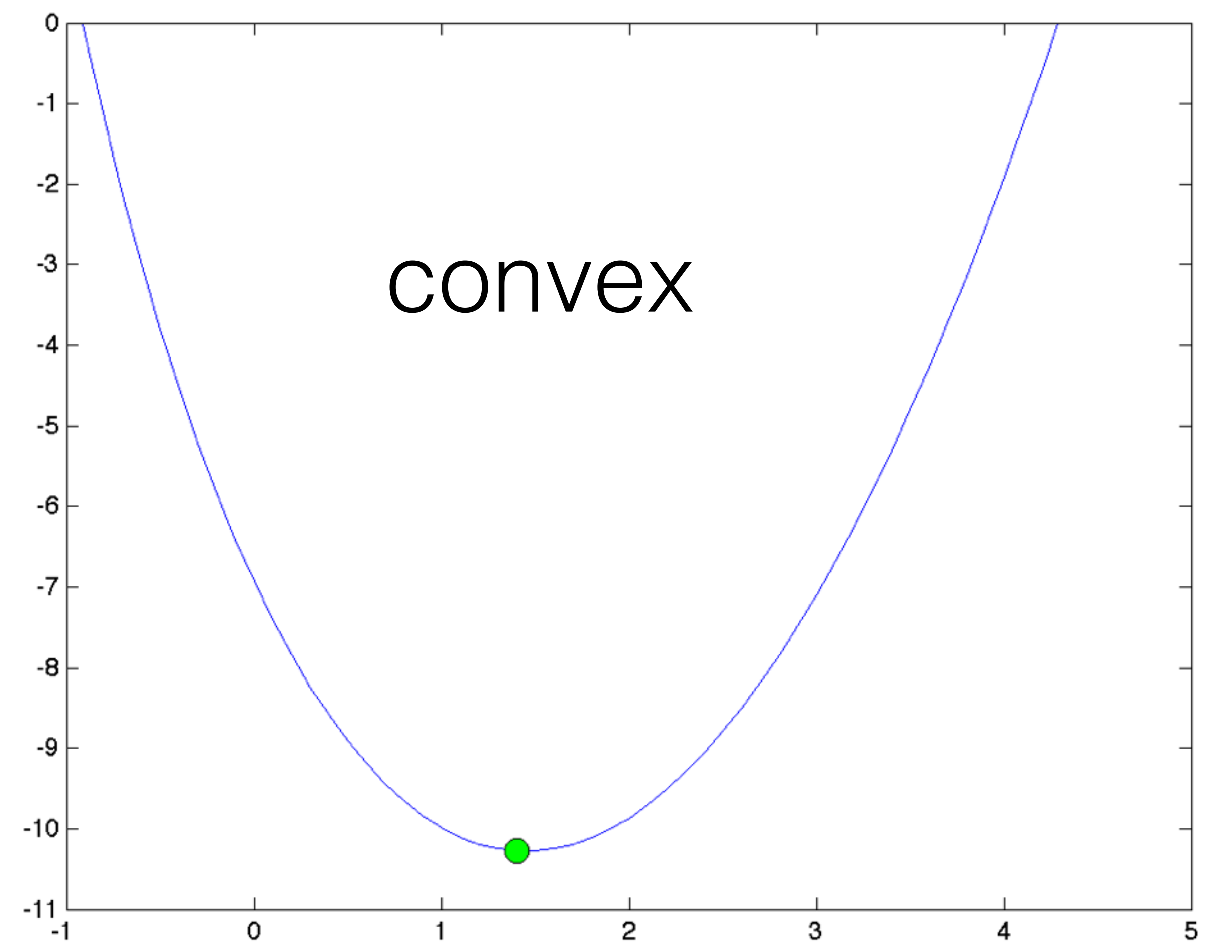
- ((x-mu)**2).sum()   works for vectors

something like: -2X.dot(Mu) + (X**2).sum(2) + (Mu**2).sum(2) ?   How about matrices?

Things you need to do in assignment 3

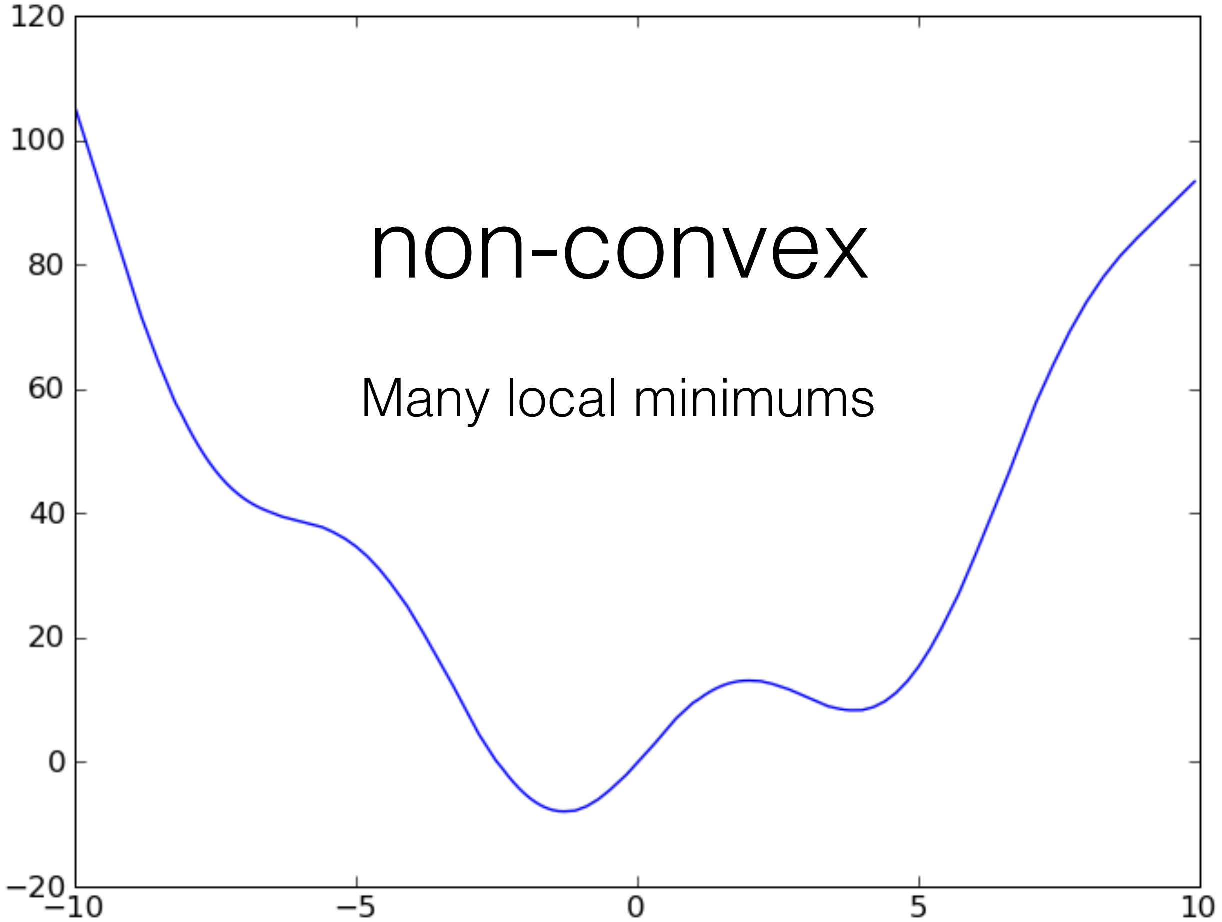
- Part 2
  - Convexity

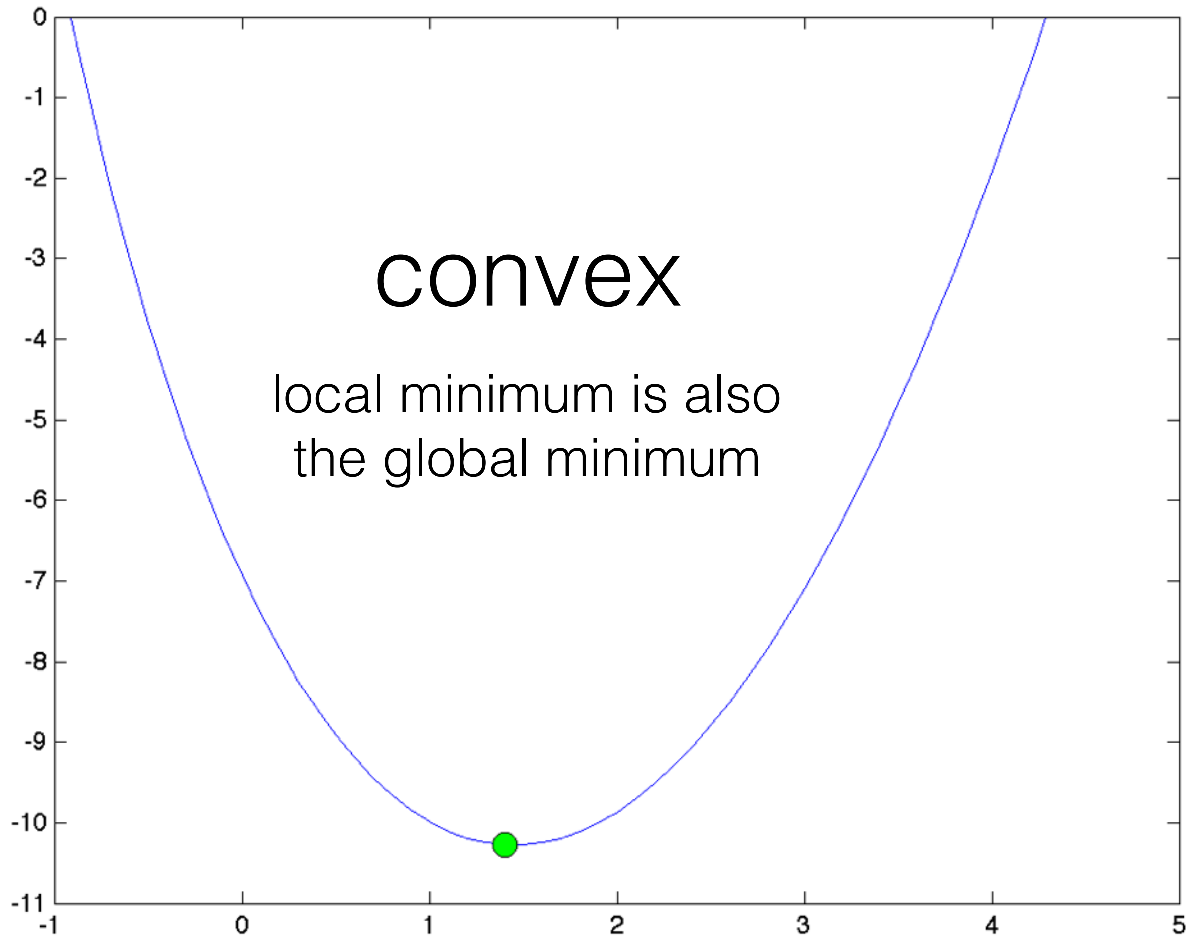




Things you need to do in assignment 3

- Part 2
  - Convexity





Things you need to do in assignment 3

- Part 2
  - Convexity
  - Code up learning. It can be done in just a few simple lines of code.

Due: Tuesday, March. 28

at midnight