281 Live Session

Week 10 — 2023/3/15

Agenda

- Overview
  - Total Least Squares
  - Gradient Descent
  - Clustering
- Exercise: Least Squares
- Assignment 6
- Group updates

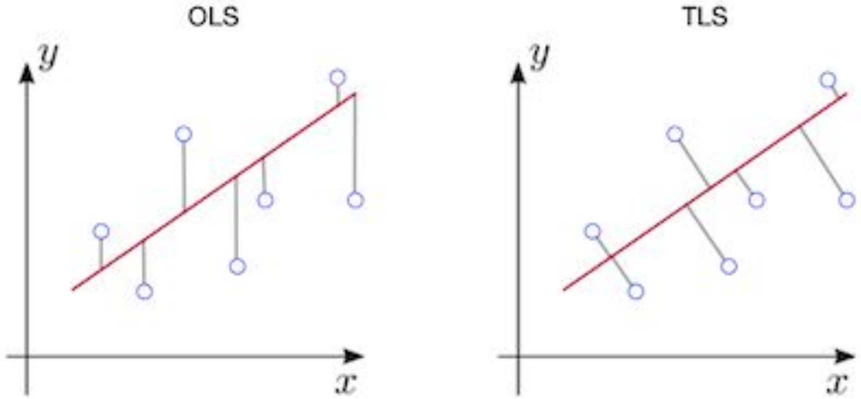
Intuition Goals

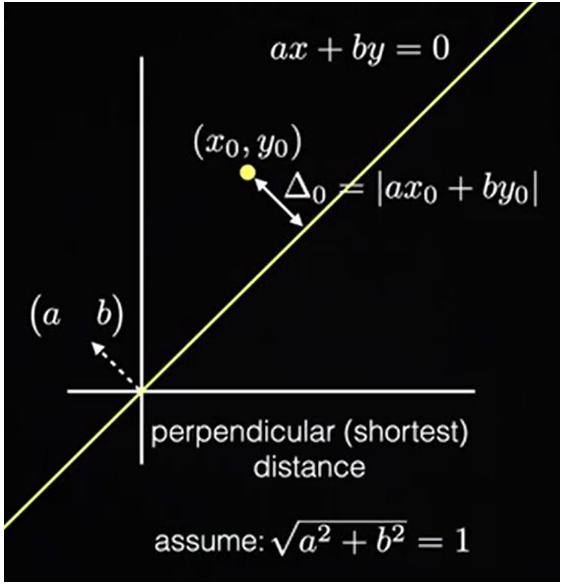
- Toolbox
- Course Overview
- Assessments
- 1: Perspective Projection
- 2: Image Formation
- 3: Image Artifacts
- 4: Convolution
- 5: Fourier
- 6: Pyramids, Edges, and Features
- 7: Image Analysis
- 8: Least-Squares
- 9: Total and Iterative Least-Squares
- 10: Clustering
- 11: Dimensionality Reduction
- 12: Linear Classifiers
- 13: Nonlinear Classifiers

- 9.1 Video Lecture Exercises
- 9.2 Line Fitting (ax + by = 0)
- 9.3 Line Fitting, Solve for a and b (ax + by = 0)
- 9.4 Total Least-Squares, Implemented
- 9.5 Least-Squares, Review
- 9.6 Steepest Descent, Intuition
- 9.7 Quadratic Form
- 9.8 Steeped Descent, Step Direction, and Size
- 9.9 Steepest Descent, Implementation
- 9.10 Conjugate Gradient Descent
- 9.11 Gradient Descent

- Least squares vs total least squares
- Gradient descent methods, strengths, and limitations
- How to choose an optimization method

Total Least Squares





Source: towardsdatascience.com

Total Least Squares

$$\begin{pmatrix} \Delta_1 \\ \vdots \\ \Delta_n \end{pmatrix} = \begin{pmatrix} x_1 & y_1 \\ \vdots & \vdots \\ x_n & y_n \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

$$\vec{\Delta} = X\vec{u}$$

$$E(\vec{u}) = \|X\vec{u}\|^2$$

$$\frac{dE}{d\vec{u}} = 2X^T(X\vec{u}) = \vec{0}$$

$$\vec{u} = (X^TX)^{-1}\vec{0}$$

$$\vec{u} = \vec{0}\ ?$$

$$E(\vec{u}, \lambda) = \|X\vec{u}\|^2 + \lambda(\|\vec{u}\|^2 - 1)$$

$$\frac{dE}{d\vec{u}} = 2X^T(X\vec{u}) + 2\lambda\vec{u} = 0$$

$$(X^TX)\vec{u} = -\lambda\vec{u}$$

$\vec{u}$ is an eigenvector of (X^TX)

$$\frac{dE}{d\lambda} = \vec{u}^T\vec{u} - 1 = 0$$

$$\vec{u}^T\vec{u} = 1$$

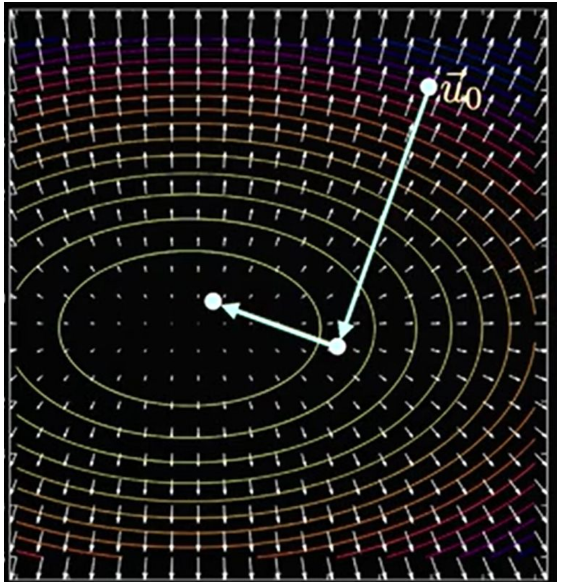
$$\vec{u}^T(X^TX)\vec{u} = -\lambda\vec{u}^T\vec{u}$$

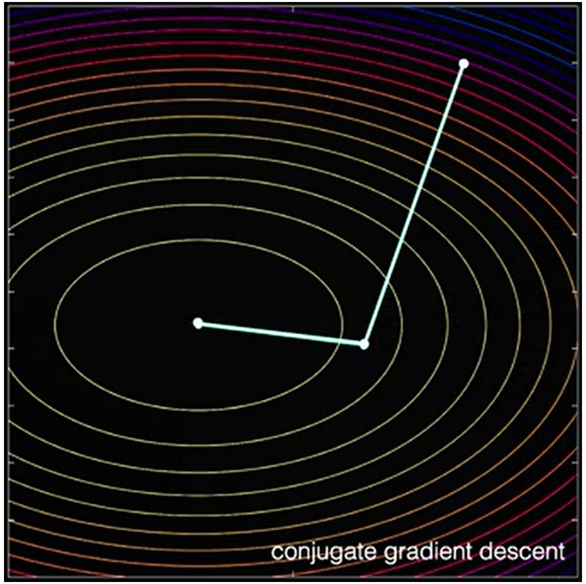
$$(\vec{u}^TX^T)(X\vec{u}) = -\lambda$$

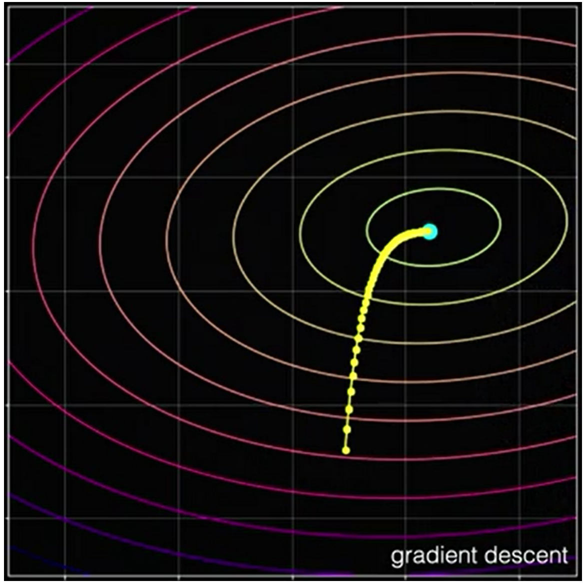
$$\|X\vec{u}\|^2 = -\lambda$$

$\vec{u}$ is the minimal-eigenvalue eigenvector of (X^TX)

Gradient Descent





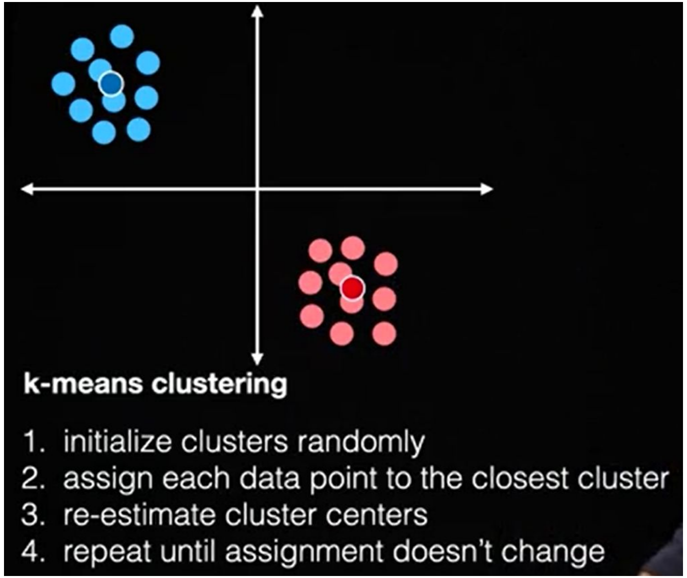


least-squares
steepest descent
conjugate gradient descent
gradient descent
stochastic gradient descent

Intuition Goals

- The importance of choosing good initial conditions
- Fundamental difference between k-means and EM
- What are these methods used for?
- What is convergence? Why might a model not converge?
- Why might a model converge to the wrong answer?

k-means/EM

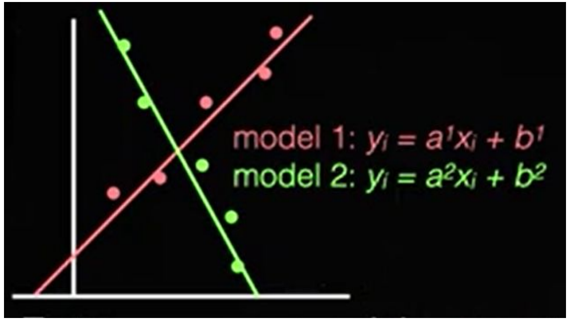


Expectation/Maximization (EM) is a two-step algorithm that iteratively estimates model parameters and model assignment

E-step: assume model parameters are known, and compute probability that each data point (x_i, y_i) belongs to each model ($k = 1,2$).

M-step: re-estimate model parameters for each model ($k = 1,2$) using probability of model assignment.

k-means/EM



$$E(a^k, b^k) = \sum_{i=1}^{n} (w_i^k (a^k x_i + b^k - y_i))^2$$

$$E(\vec{m}^k) = \left\| \begin{pmatrix} w_1^k & 0 & \cdots & 0 \\ 0 & w_2^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n^k \end{pmatrix} \left[\begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} \begin{pmatrix} a^k \\ b^k \end{pmatrix} - \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \right] \right\|^2$$

$$E(\vec{m}^k) = \| W^k (X \vec{m}^k - \vec{y}) \|^2$$

$$\boxed{\vec{m}^k = (X^T (W^k)^2 X)^{-1} X^T (W^k)^2 \vec{y}}$$

For each data point *i*, and for each model *k*, compute residual:

$$r_i^k = |a^k x_i + b^k - y_i|, \quad k = 1, 2$$

Given this residual for each model, what is the *probability* that (x_i, y_i) belongs to model *k* ?

$$\boxed{P(i \in M_k | r_i^k) = \frac{\overbrace{P(r_i^k | i \in M_k)}^{\text{likelihood}} \overbrace{P(i \in M_k)}^{\text{prior}}}{\underbrace{P(r_i^k)}_{\text{evidence}}}}$$

Group Exercise – Least Squares

Assignment 6: Lighting Estimation

Exposing Digital Forgeries by Detecting Inconsistencies in Lighting

Micah K. Johnson
Department of Computer Science
Dartmouth College
Hanover, NH 03755
kimo@cs.dartmouth.edu

Hany Farid
Department of Computer Science
Dartmouth College
Hanover, NH 03755
farid@cs.dartmouth.edu

ABSTRACT

When creating a digital composite of, for example, two people standing side-by-side, it is often difficult to match the lighting conditions from the individual photographs. Lighting inconsistencies can therefore be a useful tool for revealing traces of digital tampering. Borrowing and extending tools from the field of computer vision, we describe how the direction of a point light source can be estimated from only a single image. We show the efficacy of this approach in real-world settings.

Categories and Subject Descriptors

I.4 [**Image Processing**]: Miscellaneous

Keywords

Digital Tampering, Digital Forensics

1. INTRODUCTION

Consider the creation of a forgery showing two movie

Figure 1: A digital composite of movie stars Cher and Brad Pitt. Note that Cher was originally photographed with a fairly diffuse non-directional light source, whereas Brad Pitt was photographed with a directional light positioned to his left.

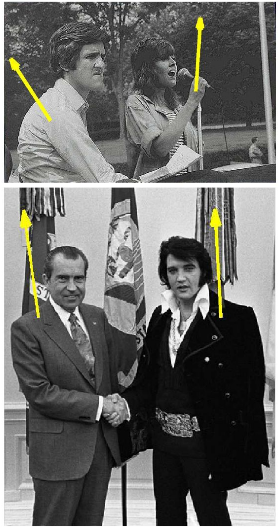
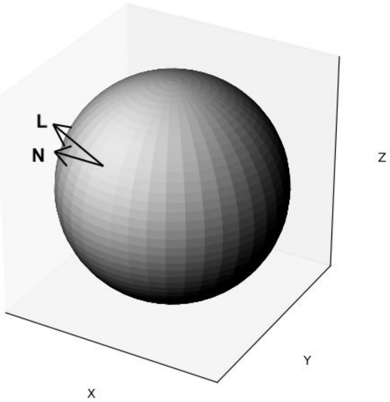


Figure 3: Shown above is a known forgery of John Kerry and Jane Fonda sharing a stage at an anti-war rally. The estimated light direction for Kerry is 123°, while the direction for Fonda is 86°. Shown below is an authentic image of Richard Nixon and Elvis Presley. The estimated directions for Nixon and Presley are 98° and 93°.



Johnson, Micah K., and Hany Farid. "Exposing digital forgeries by detecting inconsistencies in lighting." Proceedings of the 7th workshop on Multimedia and security. 2005.

Project Updates

Today: 2 minute intro

Upcoming ToDo's

Start Assignment 6 (Due Mar 28th)

Watch Async lectures for Unit 10

Prepare 2 minute project update for next week

- Include sample images per classification category
- Also include visualizations of some example feature extraction for these images, if possible