



CS145 Discussion

Week 1

Junheng, Shengming, Yunsheng
10/05/2018



- Class logistics & Reminders
- Math review: Basics of
 - Probability
 - Linear Algebra
 - Optimization
 - Matrix Calculus
- Linear Regression

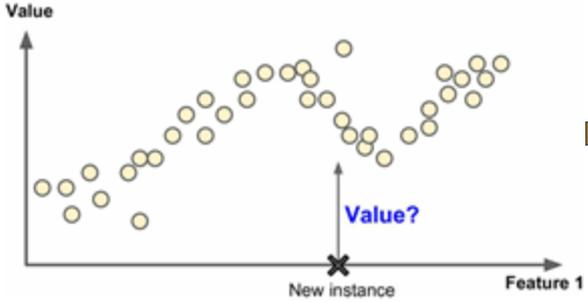
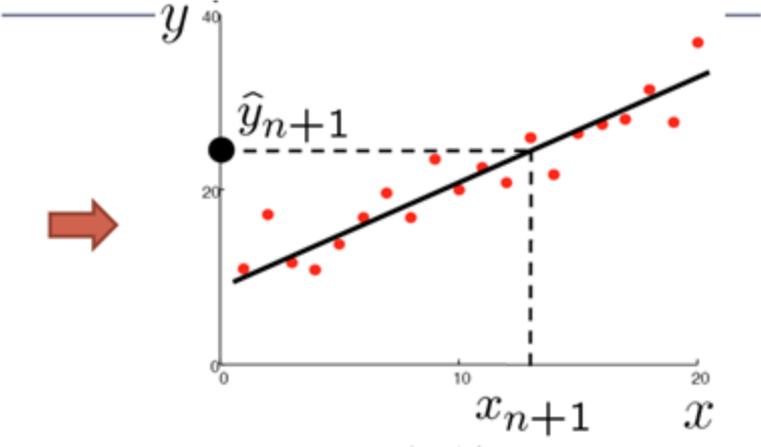
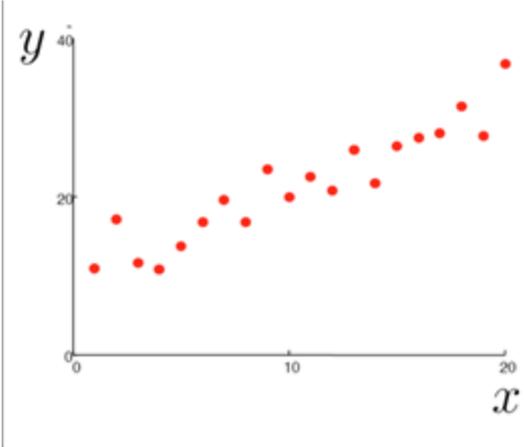


- Course Website:
 - http://web.cs.ucla.edu/~yzsun/classes/2018Fall_CS145/index.html
- Dates:
 - Project introduction on next Monday (10/08)
 - Homework 1 out on next Wednesday (10/10)
 - In-class midterm on Nov. 14 (12-2pm)
 - Final exam schedules on Dec 13 (11:30am-2:30pm)
- Update: TA office hours
 - Yizhou Sun (yzsun@cs.ucla.edu), **office hours: 3-5pm Wednesdays @BH 3531E**
 - Yunsheng Bai (yba@cs.ucla.edu), **office hours: 1-3pm Thursdays @BH 3256S**
 - Junheng Hao (haojh.ucla@gmail.com), **office hours: 1-3pm Tuesdays @BH 3256S**
 - Shengming Zhang (michaelzhang@cs.ucla.edu), **office hours: 3-5pm Mondays @BH 3256S**

Math Review

1. [Probability](#)
2. [Linear Algebra](#)
3. [Optimization](#)
4. [Matrix Calculus](#)

Linear Regression



Can this be fitted using a straight line?
No!
If not, can we still use the idea of linear regression?
Yes! (sometimes called Polynomial Regression)

Linear Regression → Polynomial Regression

- Data: n independent data objects

- $y_i, i = 1, \dots, n$

- $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T, i = 1, \dots, n$ $(x, x^2, x^3, \dots)^T$
 - A constant factor is added to model the bias term, i. e., $x_{i0} = 1$
 - New \mathbf{x} : $\mathbf{x}_i = (x_{i0}, x_{i1}, x_{i2}, \dots, x_{ip})^T$

- Model:

- y : *dependent variable*

- \mathbf{x} : *explanatory variables*

- $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$: *weight vector*

- $y = \mathbf{x}^T \boldsymbol{\beta} = \beta_0 + x_1\beta_1 + x_2\beta_2 + \dots + x_p\beta_p \rightarrow \beta_0 + x\beta_1 + x^2\beta_2 + \dots$

Matrix form of Least Square Estimation

$$J(\boldsymbol{\beta}) = (\mathbf{X}\boldsymbol{\beta} - \mathbf{y})^T (\mathbf{X}\boldsymbol{\beta} - \mathbf{y}) / 2$$

$$\begin{bmatrix} 1, x_{11} & \dots & x_{1f} & \dots & x_{1p} \\ \dots & \dots & \dots & \dots & \dots \\ 1, x_{i1} & \dots & x_{if} & \dots & x_{ip} \\ \dots & \dots & \dots & \dots & \dots \\ 1, x_{n1} & \dots & x_{nf} & \dots & x_{np} \end{bmatrix}$$

\mathbf{X} : $n \times (p + 1)$ matrix



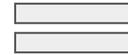
$$\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$$

$\boldsymbol{\beta}$: $(p + 1) \times 1$ matrix



$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

\mathbf{y} : $n \times 1$ matrix

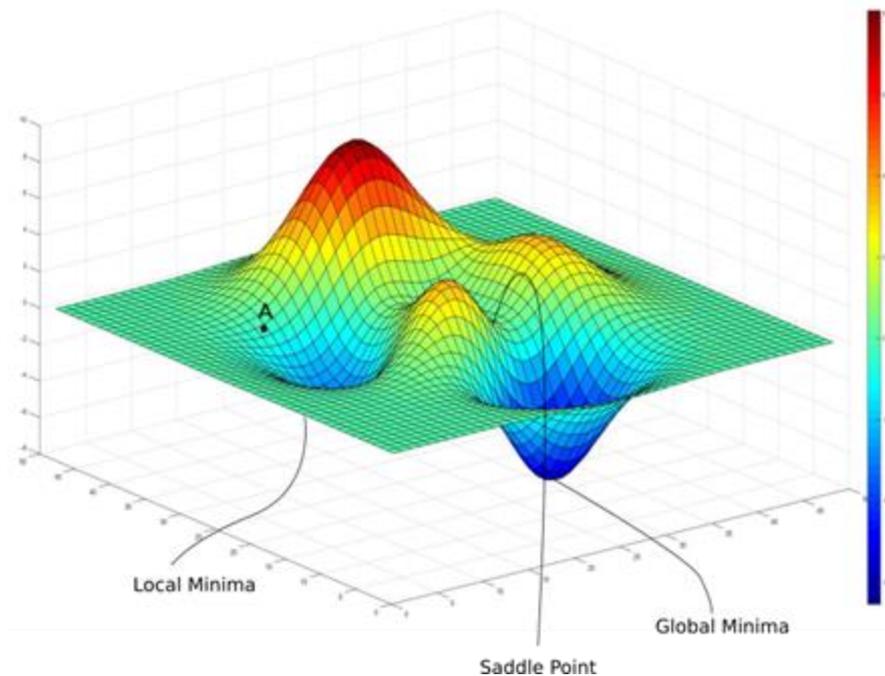
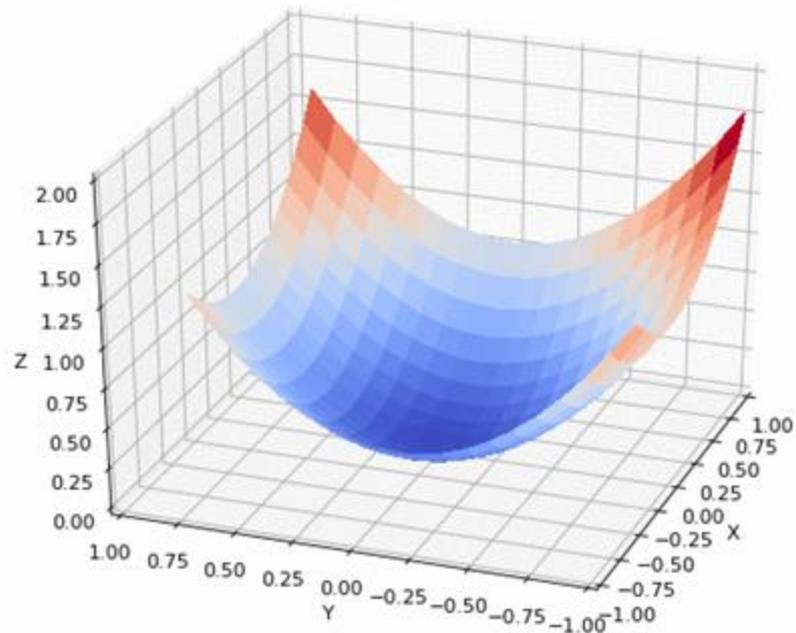


$$\begin{pmatrix} X_1 \boldsymbol{\beta} - y_1 \\ X_2 \boldsymbol{\beta} - y_2 \\ \vdots \\ X_n \boldsymbol{\beta} - y_n \end{pmatrix}$$

$\mathbf{X}\boldsymbol{\beta} - \mathbf{y}$: $n \times 1$ matrix

$$J(\boldsymbol{\beta}) = ||\mathbf{X}\boldsymbol{\beta} - \mathbf{y}||^2 / 2$$

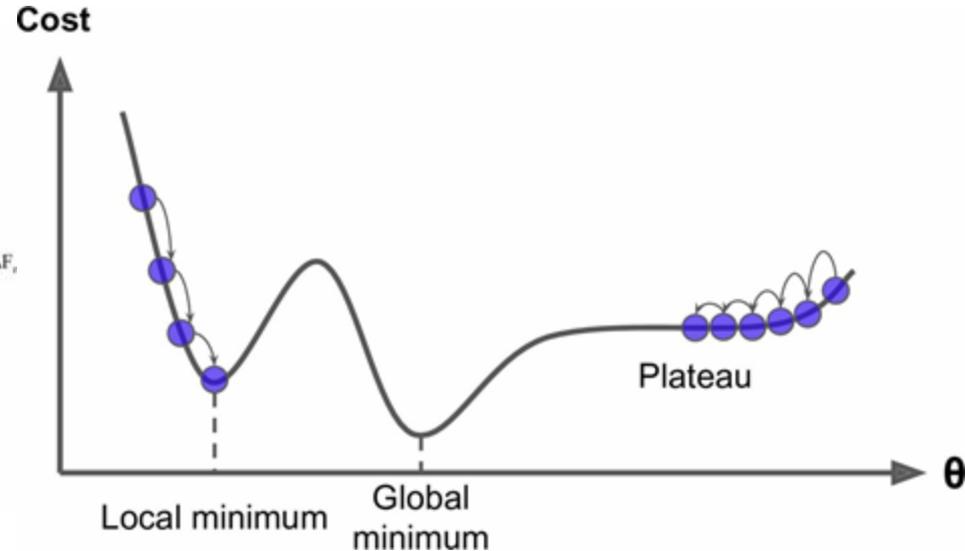
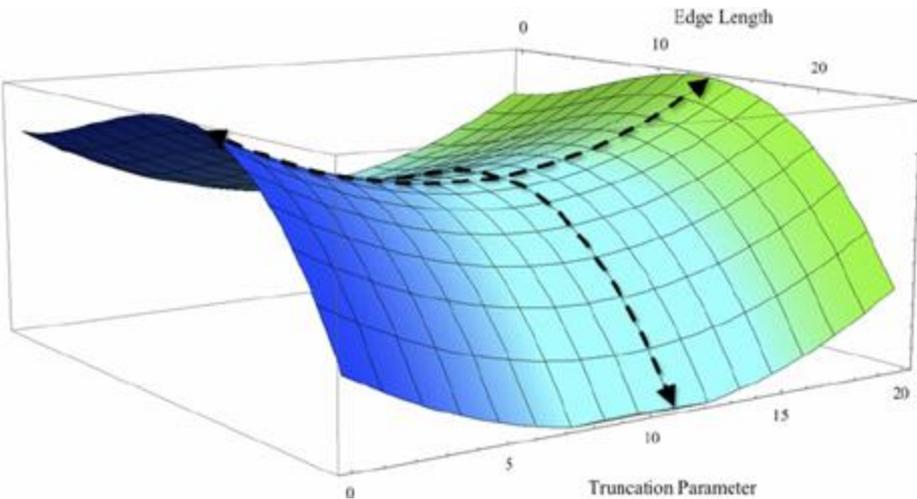
Gradient Descent



Batch vs Stochastic Gradient Descent

Why do we need Stochastic GD besides the efficiency/scalability reason?

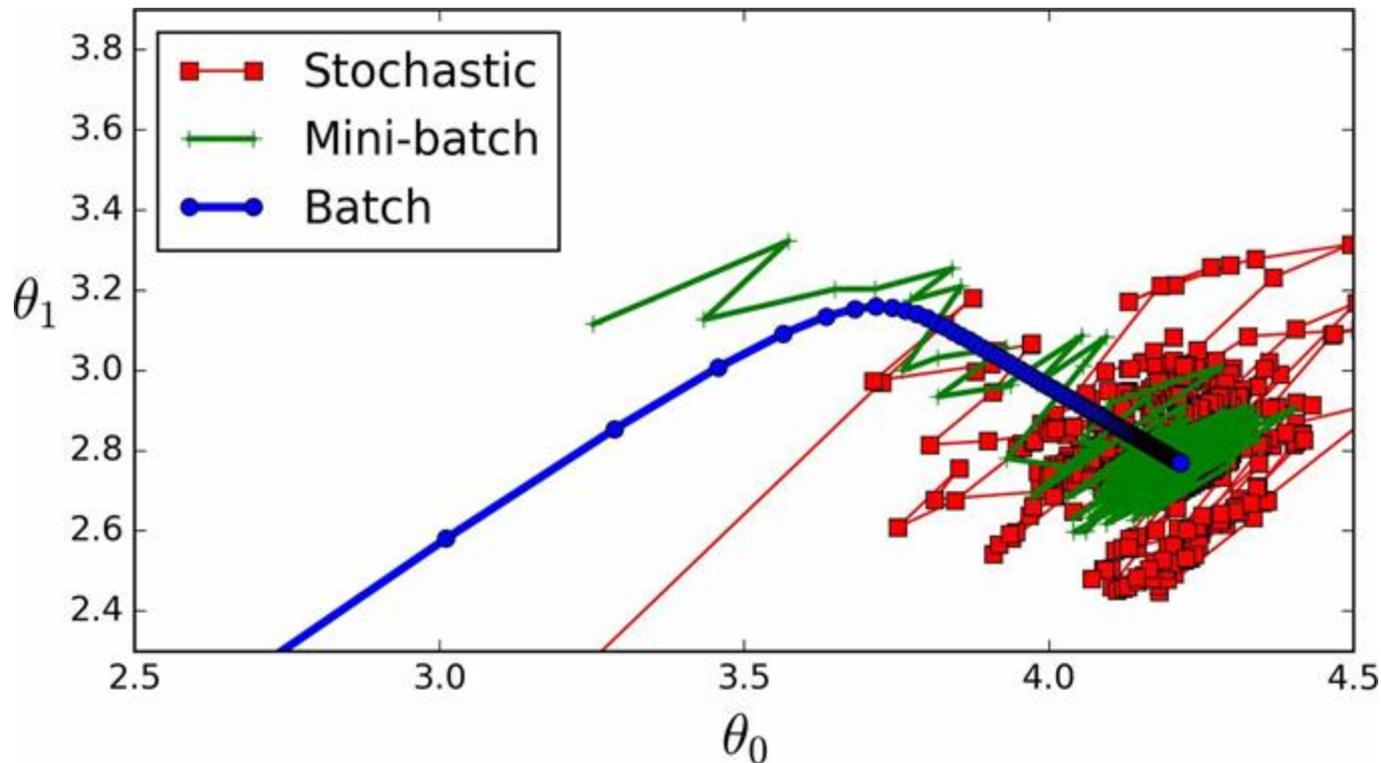
Think about saddle points and local minima.



<https://blog.paperspace.com/intro-to-optimization-in-deep-learning-gradient-descent/>

Géron, Aurélien. *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems.* " O'Reilly Media, Inc.", 2017.

Comparison



Linear Regression: When to use...

1. Closed form solution? $\Rightarrow \hat{\beta} = (X^T X)^{-1} X^T y$
2. Gradient descent
 - a. Batch GD?
 - b. Stochastic GD?
 - c. Mini-batch GD?

Feature Extraction from Real Data

- Types of Features

- Numerical
- Categorical
 - Nominal, Binary, Ordinal

- Real data may be messy for extracting features

- Unorganized structure
- Hidden and deep information

```

"created_at": "Tue Nov 24 00:14:03 +0000 2015",
"id": 6829456888951888,
"id_str": "6829456888951888",
"text": "I know that I let you down. Is it too late now to say sorry?",
"source": "no_reply/http://twitter.com/download/iphone? ref=twifollow/twitter for iPhone/iOS",
"retweeted": false,
"reply_to_status_id": null,
"reply_to_status_id_str": null,
"reply_to_user_id": null,
"reply_to_user_id_str": null,
"reply_to_screen_name": null,
"user": {
  "id": 285176387,
  "id_str": "285176387",
  "name": "SOMA X310",
  "screen_name": "twotrip",
  "location": "Newcastle",
  "url": "http://instagram.com/sinamelin",
  "description": "I'm twitrip | gonna | x310",
  "protected": false,
  "verified": false,
  "followers_count": 332,
  "friends_count": 64,
  "listed_count": 1,
  "media_count": 842,
  "statuses_count": 1332,
  "created_at": "Sun Mar 28 17:28:57 +0000 2011",
  "time_zone": "Pacific Time US & Canada",
  "geo_enabled": true,
  "lang": "en",
  "contributors_enabled": false,
  "is_translator": false,
  "profile_background_color": "000000",
  "profile_background_image_url": "https://abs.twimg.com/images/themes/theme1/bg.gif",
  "profile_background_image_url_https": "https://abs.twimg.com/images/themes/theme1/bg.gif",
  "profile_background_tile": false,
  "profile_link_color": "0000FF",
  "profile_sidebar_border_color": "000000",
  "profile_sidebar_fill_color": "000000",
  "profile_text_color": "000000",
  "profile_use_background_image": true,
  "profile_image_url": "http://pbs.twimg.com/profile_images/668769724526888/rfW2xjt_normal.jpg",
  "profile_image_url_https": "https://pbs.twimg.com/profile_images/668769724526888/rfW2xjt_normal.jpg",
  "default_profile": false,
  "default_profile_image": false,
  "following": null,
  "following_request_sent": null,
  "notifications": null
},
"geo": null,
"coordinates": null,
"place": {
  "id": "502c7f6b98f369",
  "url": "https://api.twitter.com/1/geo/id/502c7f6b98f369.json",
  "place_type": "city",
  "name": "San Jose",
  "full_name": "San Jose, CA",
  "country_code": "US",
  "country": "United States",
  "bounding_box": {
    "type": "Polygon",
    "coordinates": [
      [
        [
          -122.839351,
          37.583164
        ],
        [
          -122.839351,
          37.689154
        ],
        [
          -122.71215,
          37.689154
        ],
        [
          -122.71215,
          37.583164
        ],
        [
          -122.839351,
          37.583164
        ]
      ]
    ]
  }
},
"contributors": {},
"retweets": {
  "count": null,
  "is_quote_status": false,
  "retweeted": false,
  "retweet_count": 0,
  "favorite_count": 0,
  "entities": {
    "hashtags": [],
    "urls": [],
    "user_mentions": [],
    "symbols": []
  }
},
"favorited": false,
"retweeted": false,
"filter_level": "low",
"lang": "en",
"retweeted_me": "540032444763"
}

```

Example of a tweet data

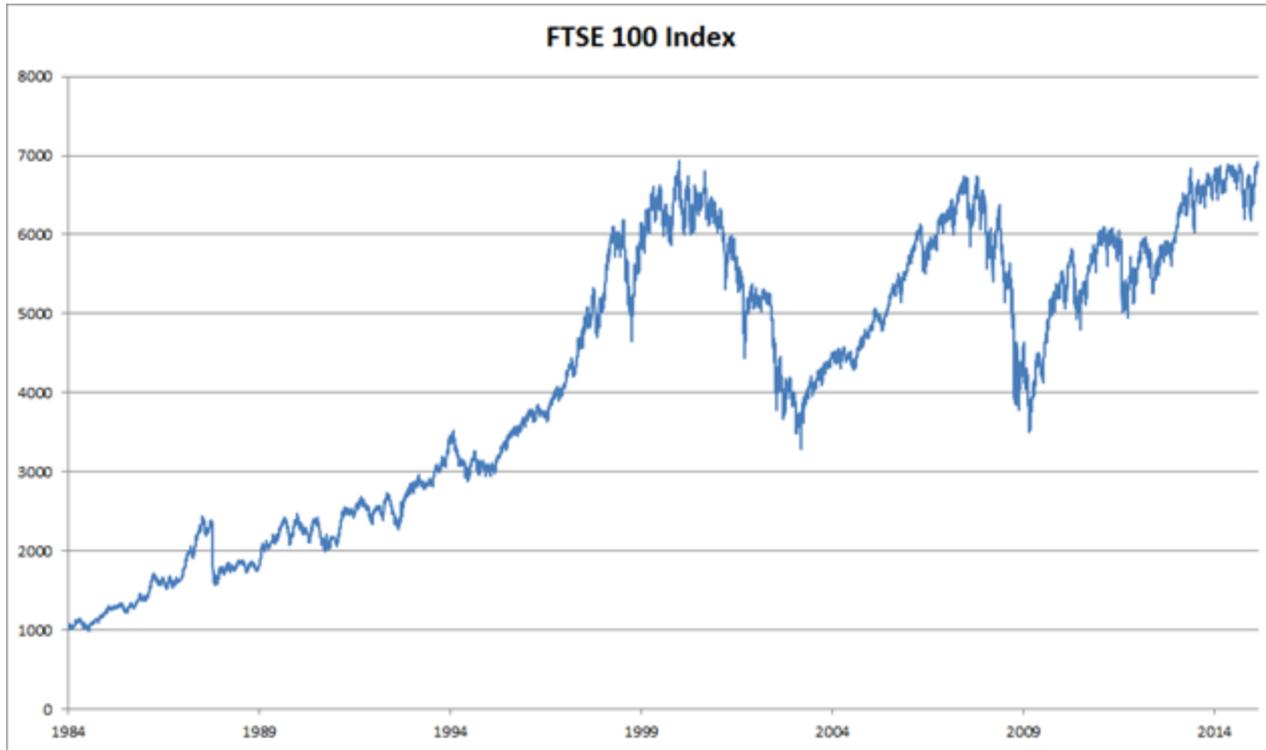
Numerical Features

- Numerical attributes
 - Raw data with numerical formats
 - E.g., numbers of friends and followers, timestamps
- Numerical statistics
 - Numerical statistics towards a characteristic
 - E.g., the length of text, the average daily number of tweets for the user
- Numerical hidden representations
 - Represent data in optimized hidden spaces
 - E.g, pLSA and LDA for text (Week 10)

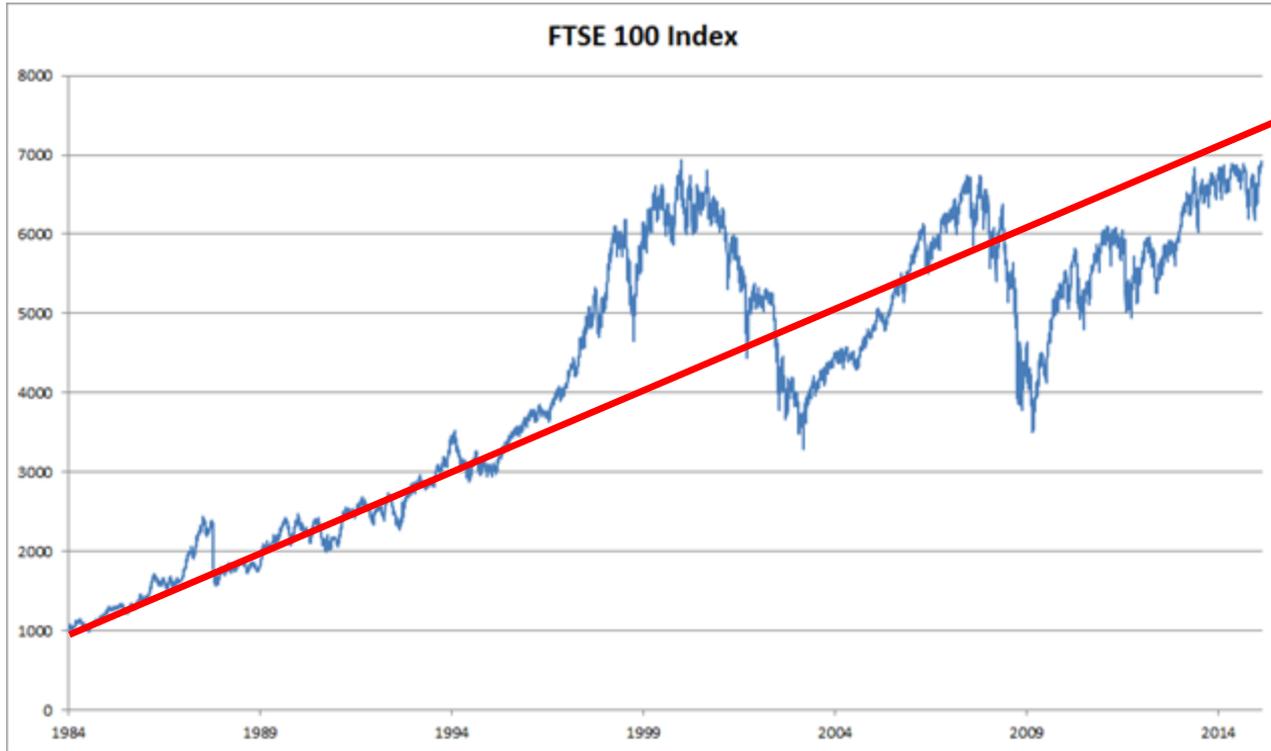
Categorical Features

- Categorical attributes
 - Raw data which originally have a set of discrete categories
 - E.g., cities of users, languages of text,
- Discretization for numerical attributes
 - Transform numerical features into categorical features
 - E.g., Morning/Afternoon/Night, Long/Short Text (more than k words?)
- Categorical statistics
 - Categorical statistics towards a characteristic
 - E.g., If the user posts more than k tweets in a week, Few/Usual/Many tweets posted in near regions

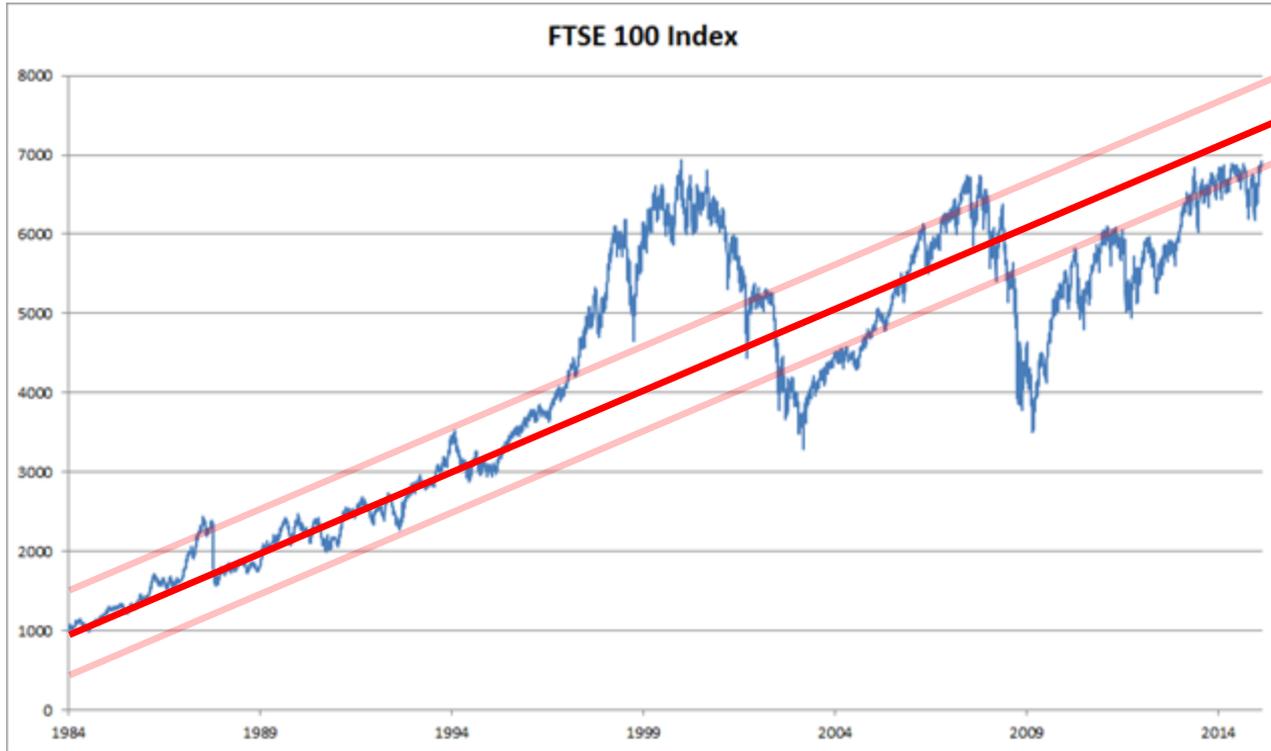
An application of linear regression: Stock Prices



An application of linear regression: Stock Prices



An application of linear regression: Stock Prices





Sample variance vs Variance

Proof & Explanation: https://en.wikipedia.org/wiki/Bessel%27s_correction

UCLA



Thank you!

Q & A