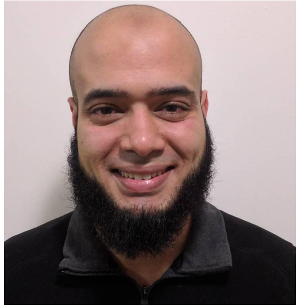
Machine Learning

ML Big Picture

Mostafa S. Ibrahim
Teaching, Training and Coaching for more than a decade!

Artificial Intelligence & Computer Vision Researcher
PhD from Simon Fraser University - Canada
Bachelor / MSc from Cairo University - Egypt
Ex-(Software Engineer / ICPC World Finalist)



Machine Learning

Approaches

- Supervised learning
- Unsupervised learning
- Reinforcement learning

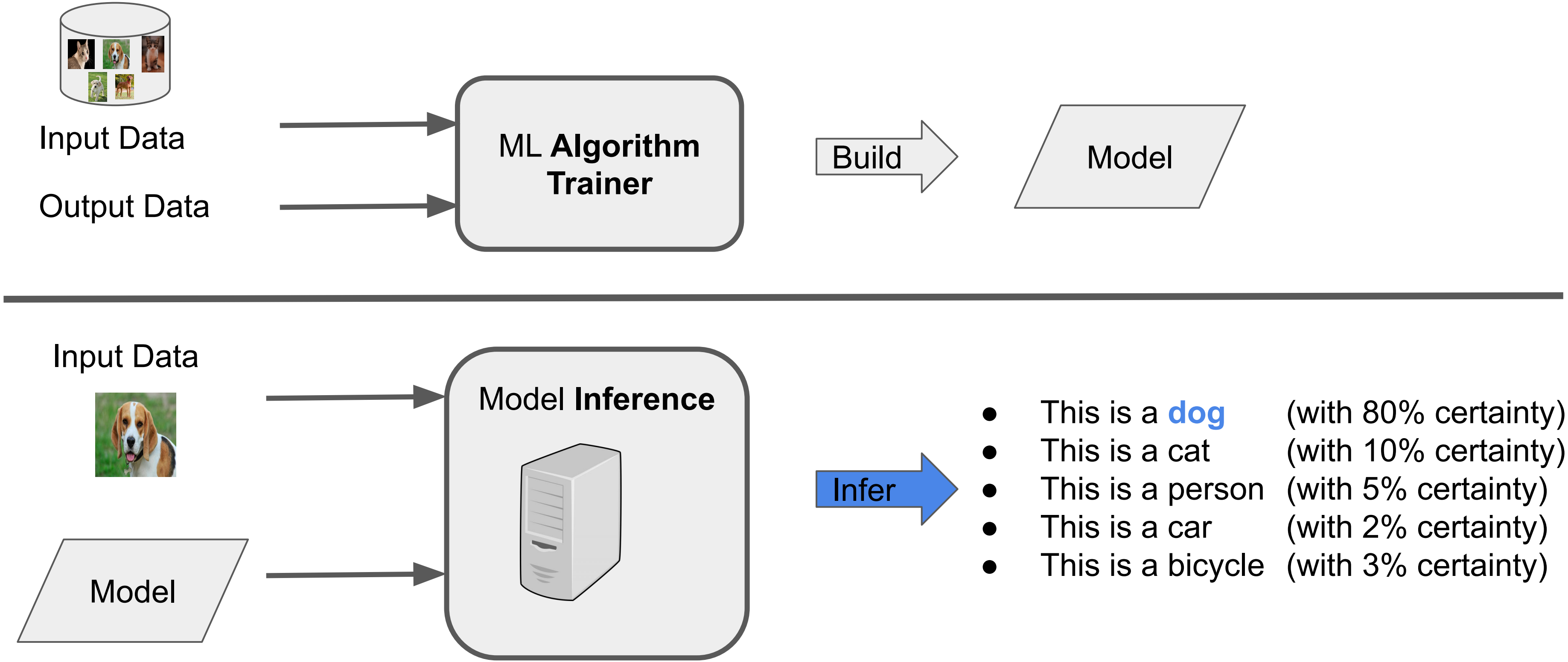
Problem Types

- Regression
- Classification
- Forecasting
- Clustering
- Recommendation

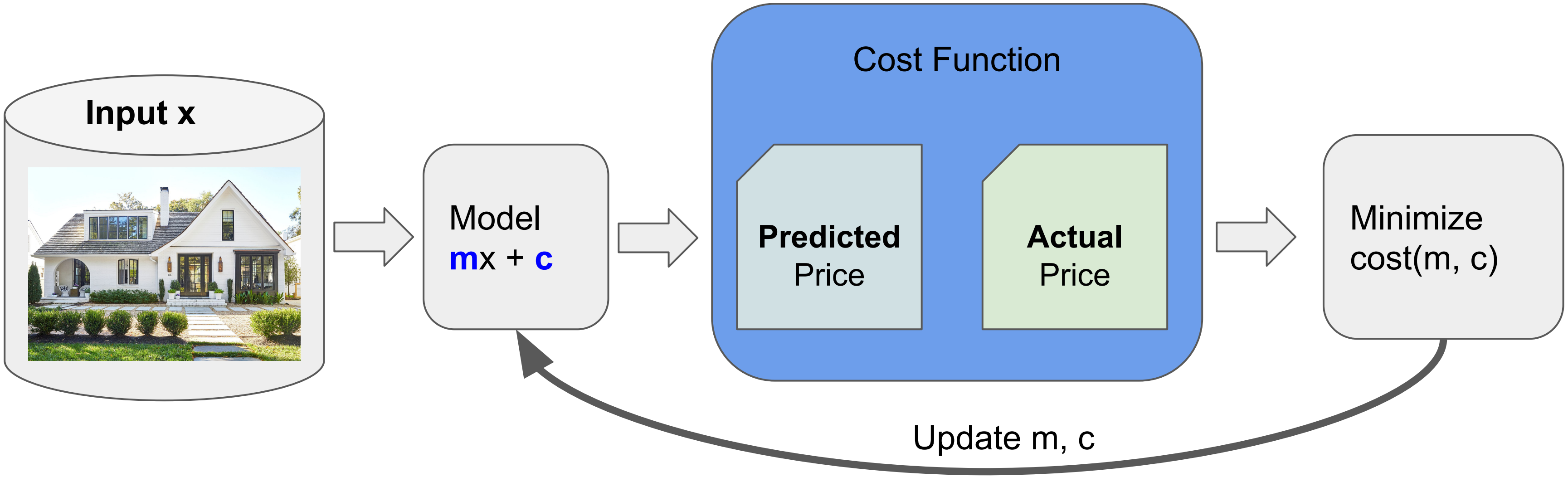
Algorithms

- Linear Regression
- Logistic regression
- Neural Network
- Deep Learning
- Tree-based Algorithms

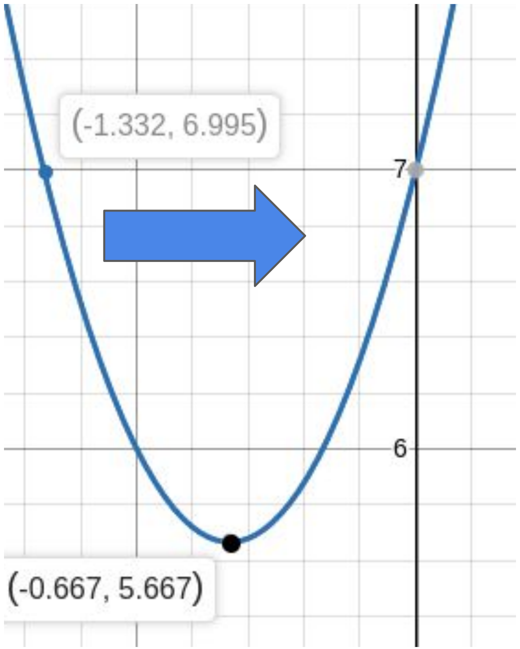
Supervised Training vs Inference (test)

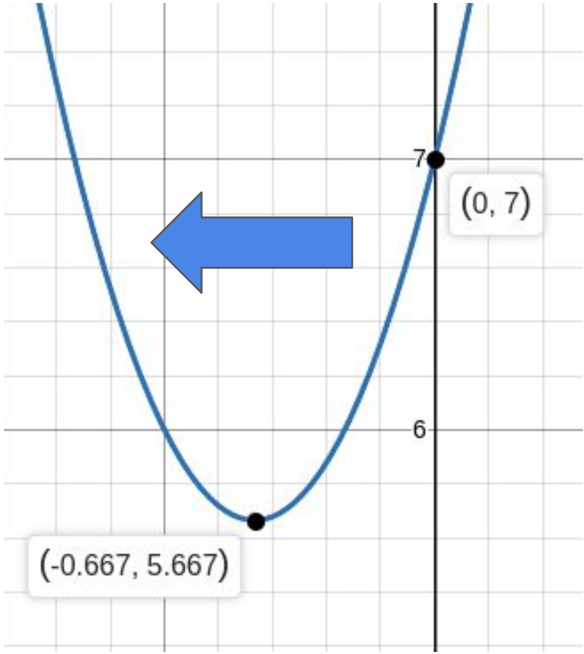


Supervised Learning



Gradient Descent: **opposite direction** of the slope





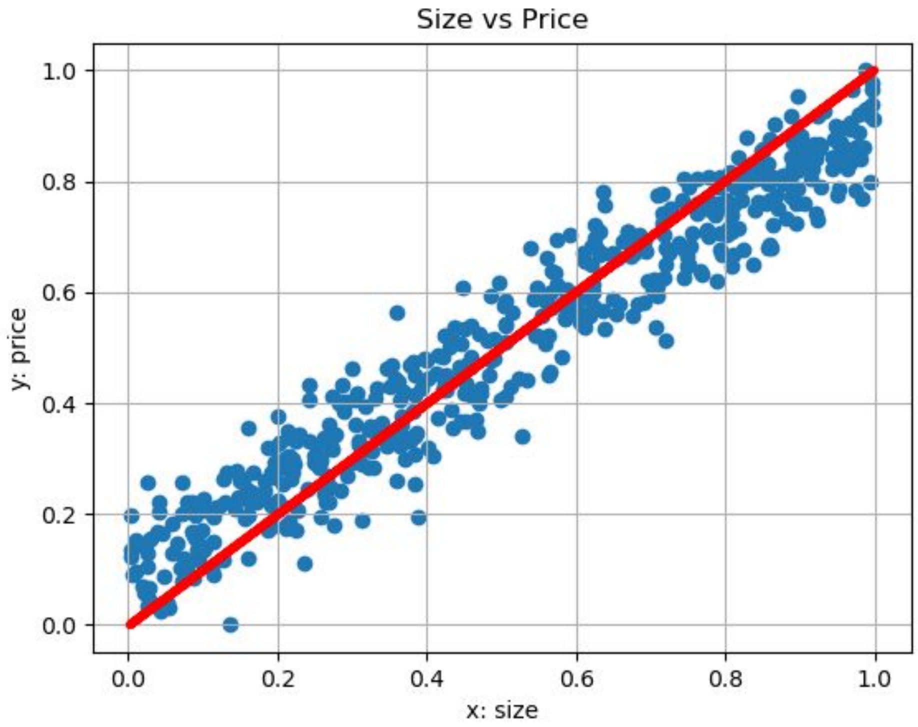

```
for iter in range(100):
    gradient = f_derivative(cur_x)
    cur_x -= gradient * step_size
```

Linear Regression

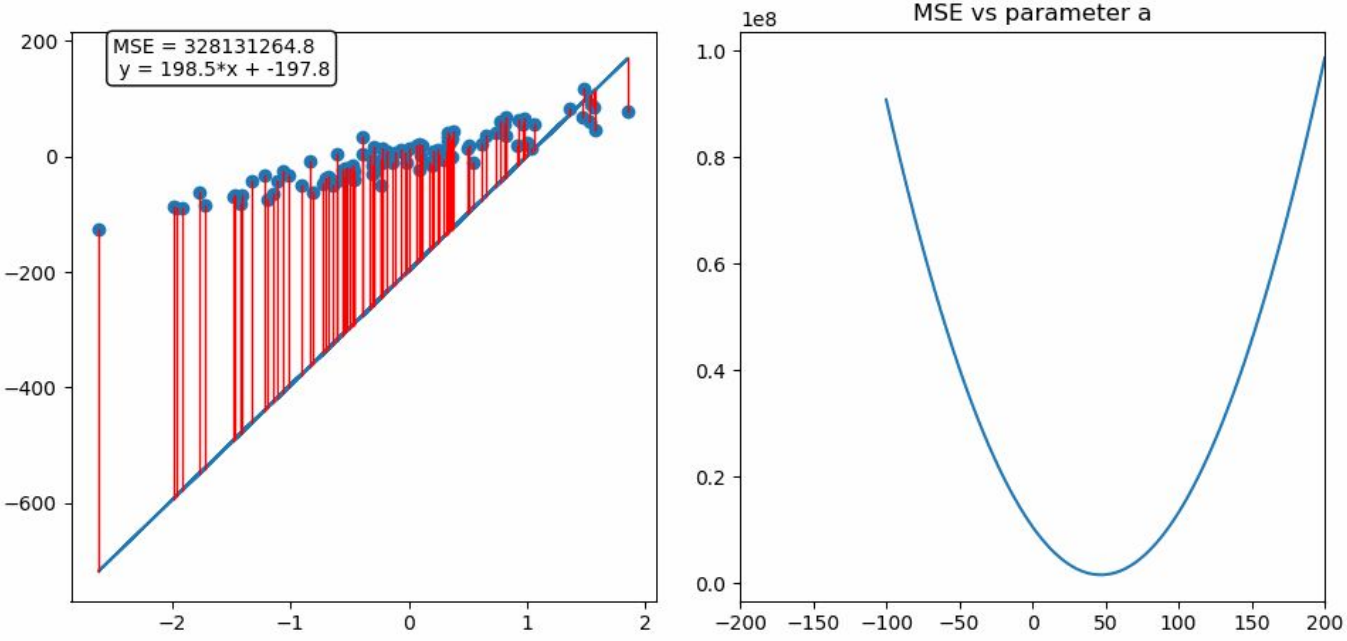
- How to find the best line (mx+c) that fits the data!
- Mean Squared Error (MSE)
  To evaluate a line against a dataset

$$\underset{m,c}{\text{minimize}}\, cost(m,c)$$

$$cost(m,c) = \frac{1}{2N}\sum_{n=1}^{N}((mx^{(n)}+c)-t^{(n)})^2$$



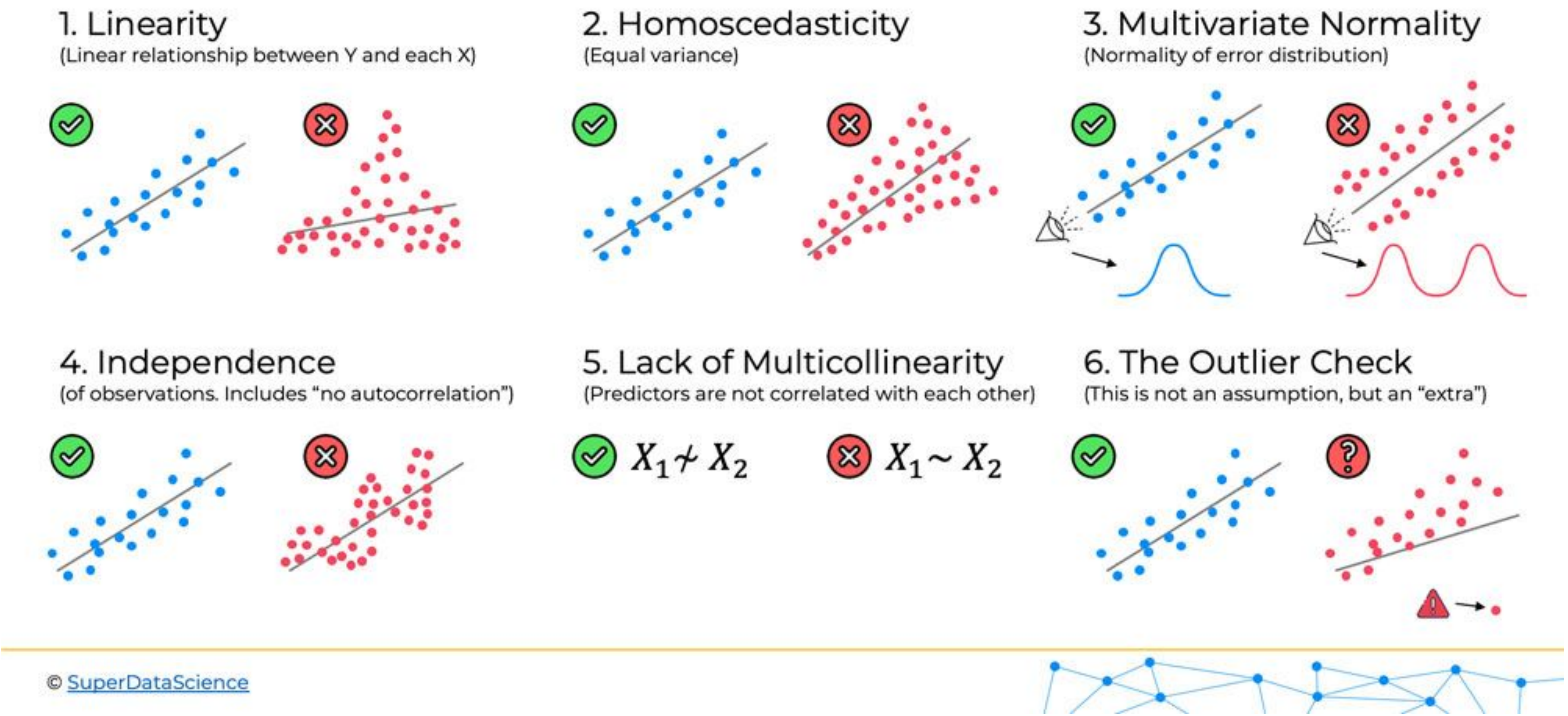
Linear Regression using Gradient Descent



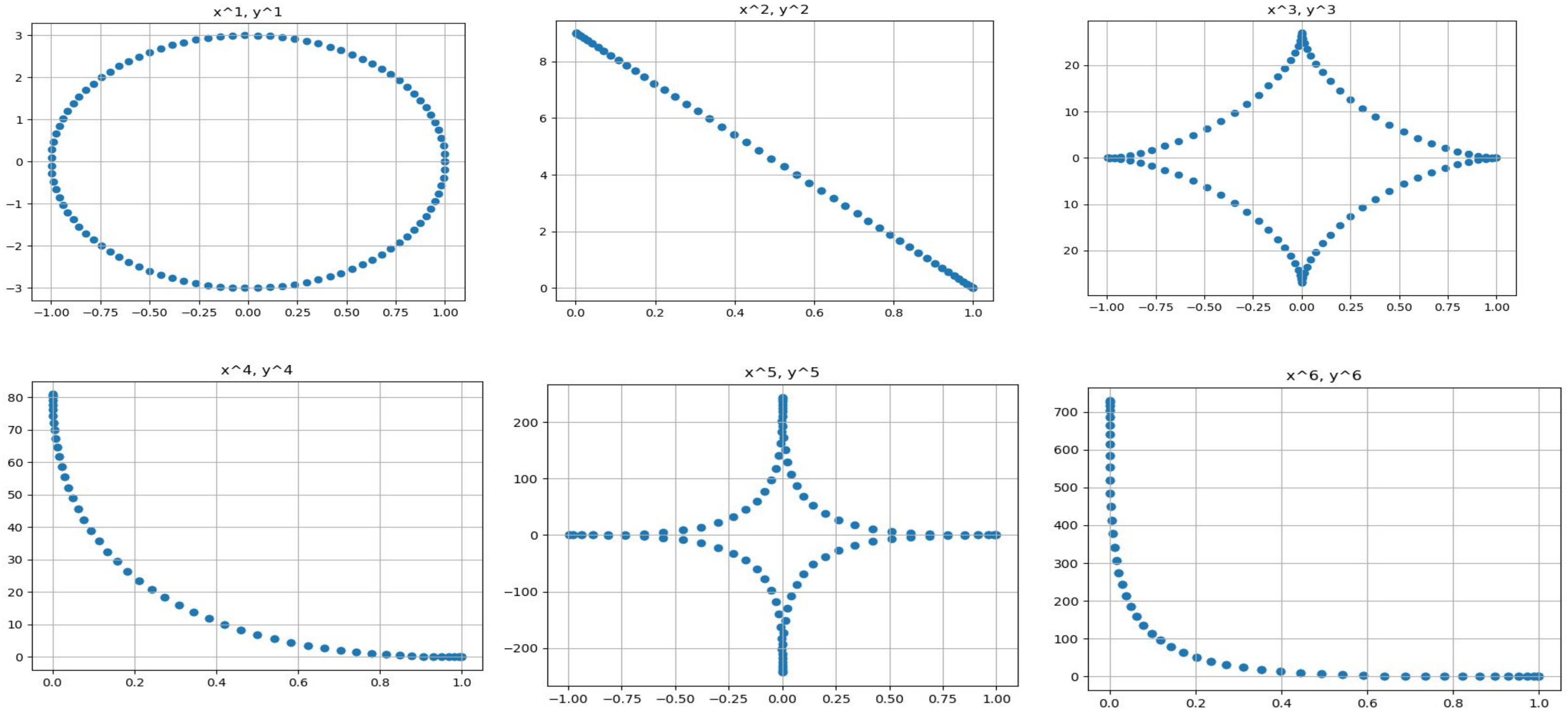
$$cost(m,c) = \frac{1}{2N}\sum_{n=1}^{N}((mx^{(n)}+c)-t^{(n)})^2$$

Img src

Linear Regression Assumptions



Space Transformation



Models

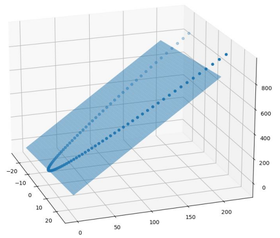
Linear Regression

model linear relationships between X and y

$$y(X^n, W) = W^T * X^n$$

Basis Regression

model non-linear relationships between X and y using a linear model (coefficients W)



Ridge Regression

Penalize with $||W||^2$ to avoid **overfitting**

$$cost(W) = \frac{1}{2N} \sum_{n=1}^{N} (y(X^n, W) - t^n)^2 + \sum_{i=1}^{M} \frac{\lambda}{2} W_i^2$$

Lasso Regression

Penalize with |W| to avoid overfitting ⇒ sparse

Select best model vs feature selection?

Modeling Concepts

Hyperparameters

- Learning Rate ∝
- Regularization lambda λ
  - Ridge / Lasso
- Grid Search / Pipeline

- K in CV-fold (5)
- Random seed?! (avoid)

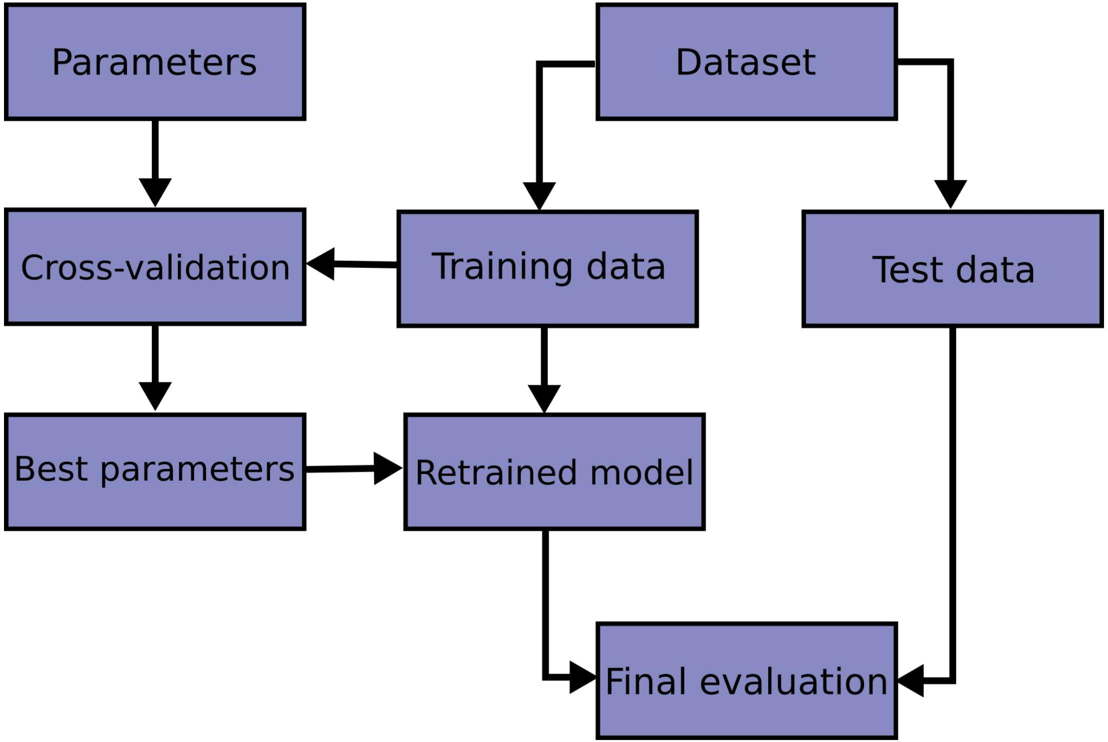
May generalize?

- Train/Val/Test Split
  - Selection bias
- Cross Validation
  - K-Fold
  - LOGOCV (groups)
  - Models: mean/**std**
- Model Selection

Fitting

- **Overfit** (low train error / high val)
  - Increase Regularization
  - Reduce Complexity
  - More Data
  - Less Features
  - Proper Training Stop
- **Underfit** (high train error)
  - Decrease Regularization
  - Increase Complexity
  - More Features
  - More training steps
- **Bias-Variance** Trade Offs
  - Bias: Due to model assumptions
  - Variance: Due to model's sensitivity to data changes
  - Practically: Test set / Regularize / Hyperparameter Tuning

Modelling Flow for small datasets



Img src

Data Wrangling (Munging)

- Data Acquisition
- Data Cleaning
  - outliers, missing data, duplicates
- Data Transformation
- Data Enrichment
- Data Integration

Data Acquisition

- **Representative** Sample
- Fine-grained vs Coarse
- Data Annotation
  - Manual - Pre-label
- Data Validation

Data Concepts

Feature Engineering

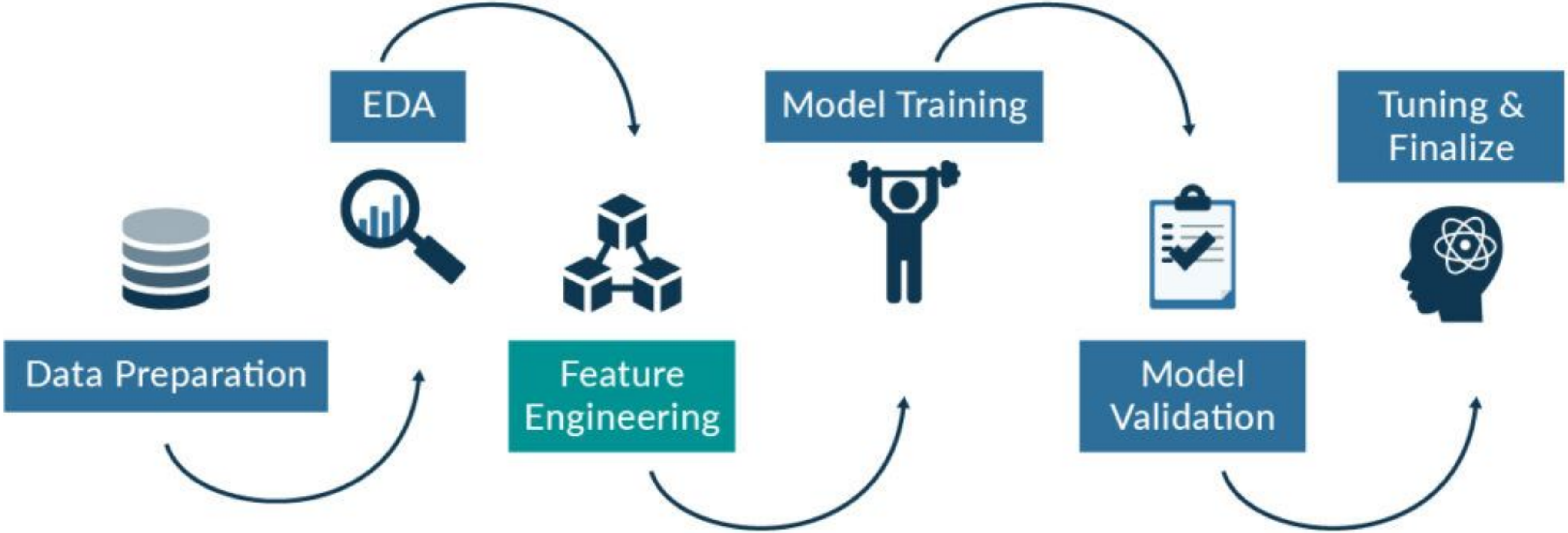
- Strings ⇒ Hash encoding / Ordinal Enc
- Integers
  - One-hot encoding
  - Binning (Discretization)
- Floating
  - Log transform for large values
  - Variance stabilizing transform
  - Scaling (minmax / standrize)

Data Enrichment

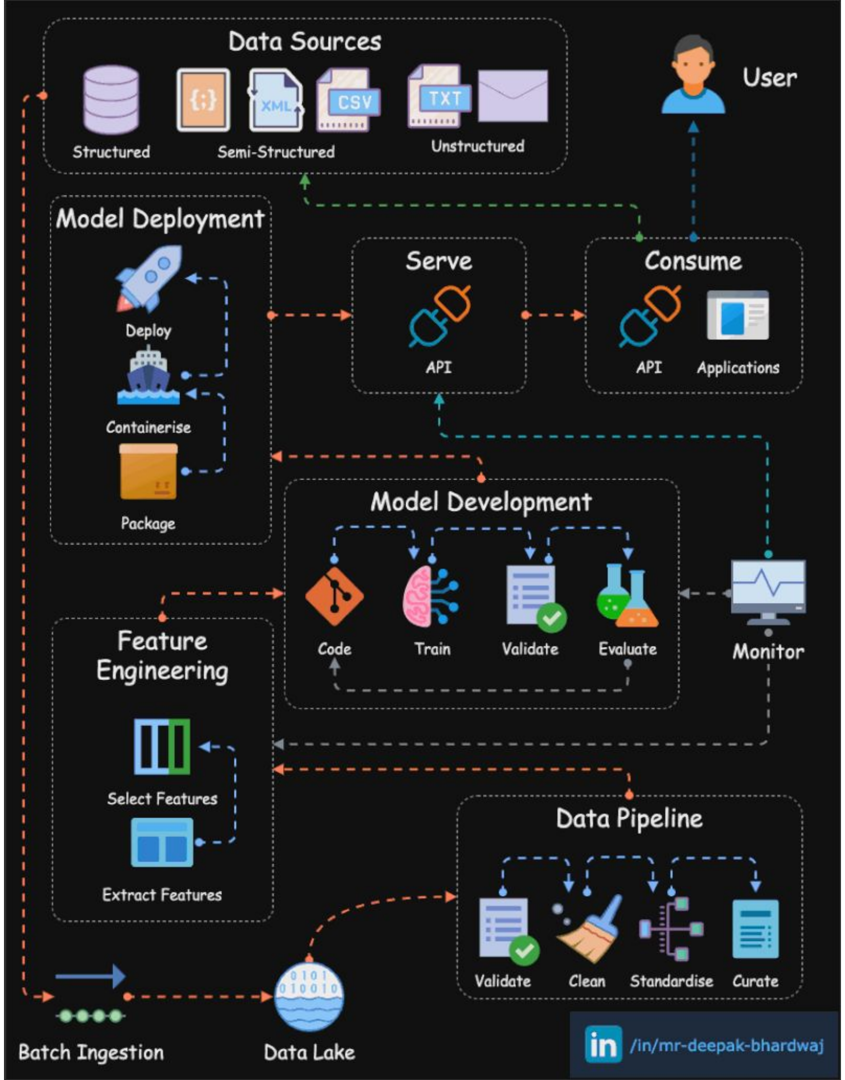
- External sources for raw data / metadata
- Invent new features / Features cross
- Data Augmentation / Synthetic Data

Issues

- Data Leakage
- Distribution Shift


EDA
Model Training
Tuning &
Finalize
Data Preparation
Feature
Engineering
Model
Validation

Img src


Data Sources
Structured
{;}
XML
CSV
Semi-Structured
TXT
Unstructured
User
Model Deployment
Deploy
Containerise
Package
Serve
API
Consume
API
Applications
Model Development
Code
Train
Validate
Evaluate
Monitor
Feature Engineering
Select Features
Extract Features
Data Pipeline
Validate
Clean
Standardise
Curate
Batch Ingestion
0101
010010
Data Lake
in /in/mr-deepak-bhardwaj