Components of Machine Learning Paradigm

Yan Yan

Review of last class

- What is learning
- What is machine learning paradigm
- What are pre-defined settings (tasks, supervision)
- What are features (discriminative properties)
- What are learning models (structures, parameter adjustment)

Today’s class includes

- Machine learning task details: classification, clustering and regression
  - Terminology
- Their connection to real world applications
  - Why we need these tools
- A showcase: how to construct a learning model
  - What can be used as features
  - Determine model structure
  - Determine model parameters

Components of machine learning paradigm

Past data: features

How to choose/generate useful features?

Learning model for bus recognition

How to determine this model?

Prediction: bus

Future data

Pre-defined settings/tasks

Components of machine learning paradigm

Past data: features

How to choose/generate useful features?

Learning model for bus recognition

How to determine this model?

Prediction: bus

Future data

Pre-defined settings/tasks

Three key components

Components of machine learning paradigm

Training data

How to choose/generate useful features?

Learning model for bus recognition

How to determine this model?

Prediction: bus

Testing data

Pre-defined settings/tasks

Three key components

Components of machine learning paradigm

Training data

How to choose/generate useful features?

Learning model for bus recognition

How to determine this model?

Prediction: bus

Testing data

Pre-defined settings/tasks

From real applications: providing demand/requirement of problems

Machine learning tasks

- Classification: traffic sign recognition

Learning model for traffic sign recognition

Machine learning tasks

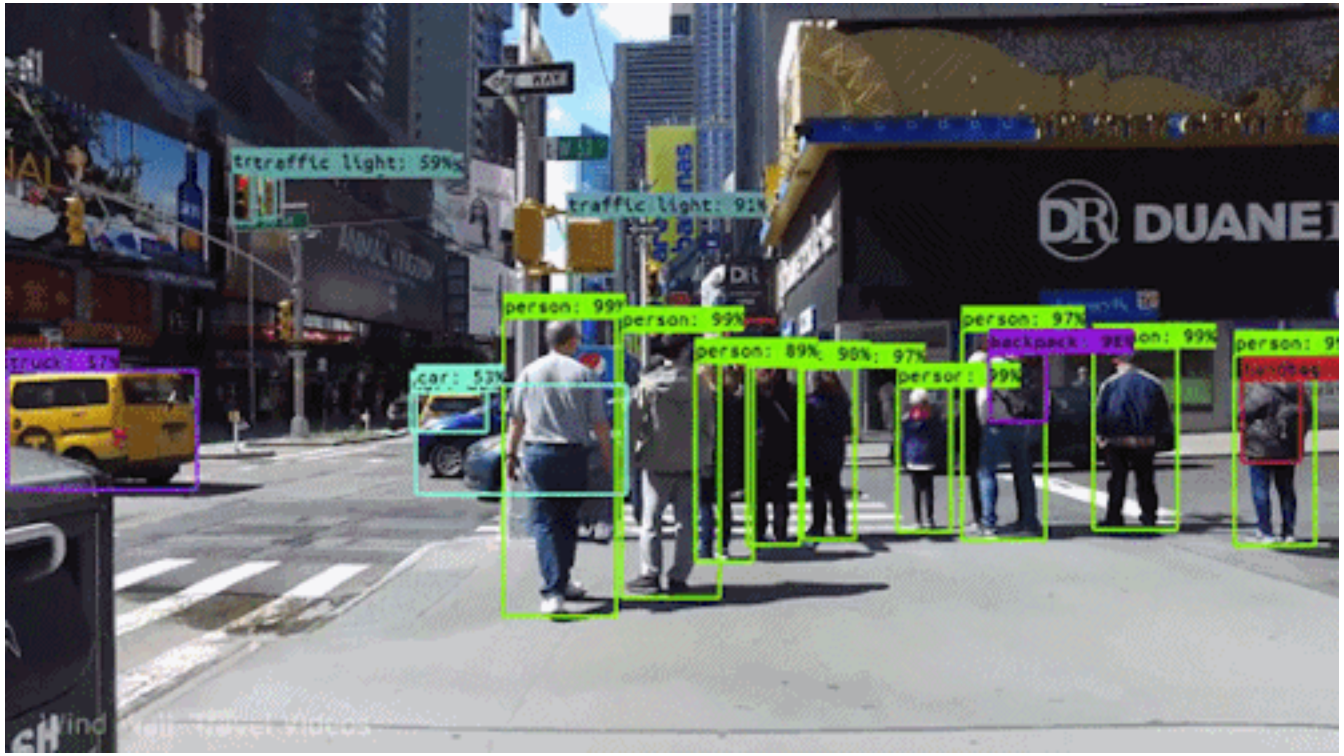
- Classification: traffic sign recognition

Learning model for traffic sign recognition

Q: How we can use it in practice?

Machine learning tasks

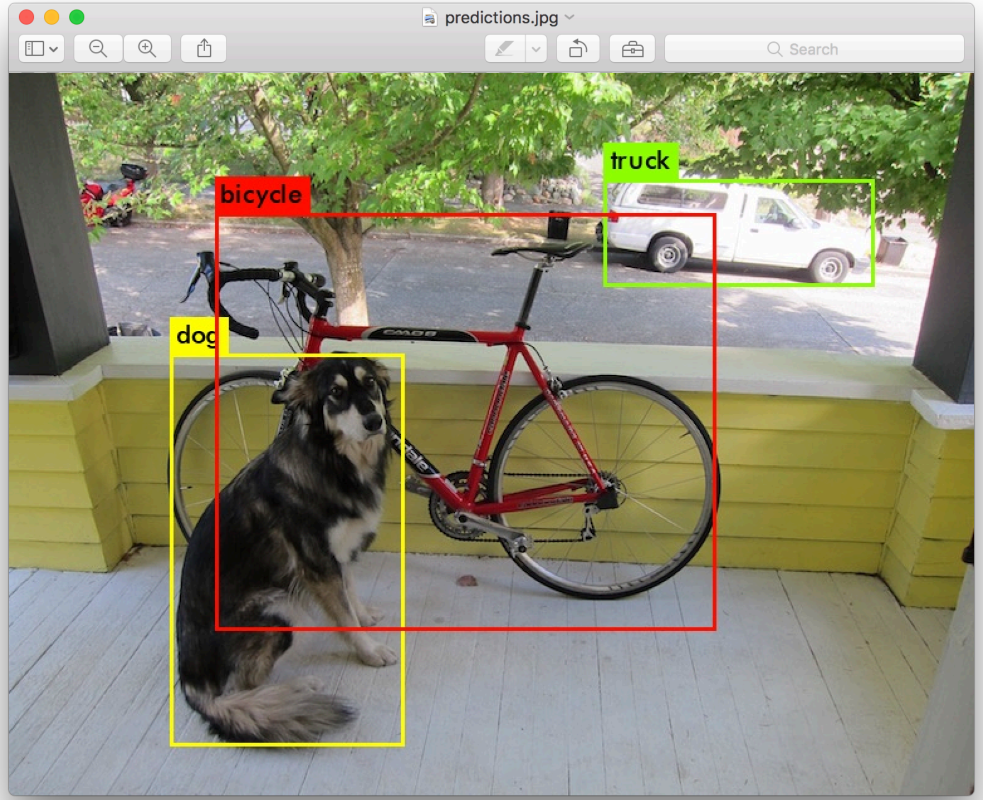
- Classification: traffic sign recognition



A use example: autonomous driving system

Machine learning tasks

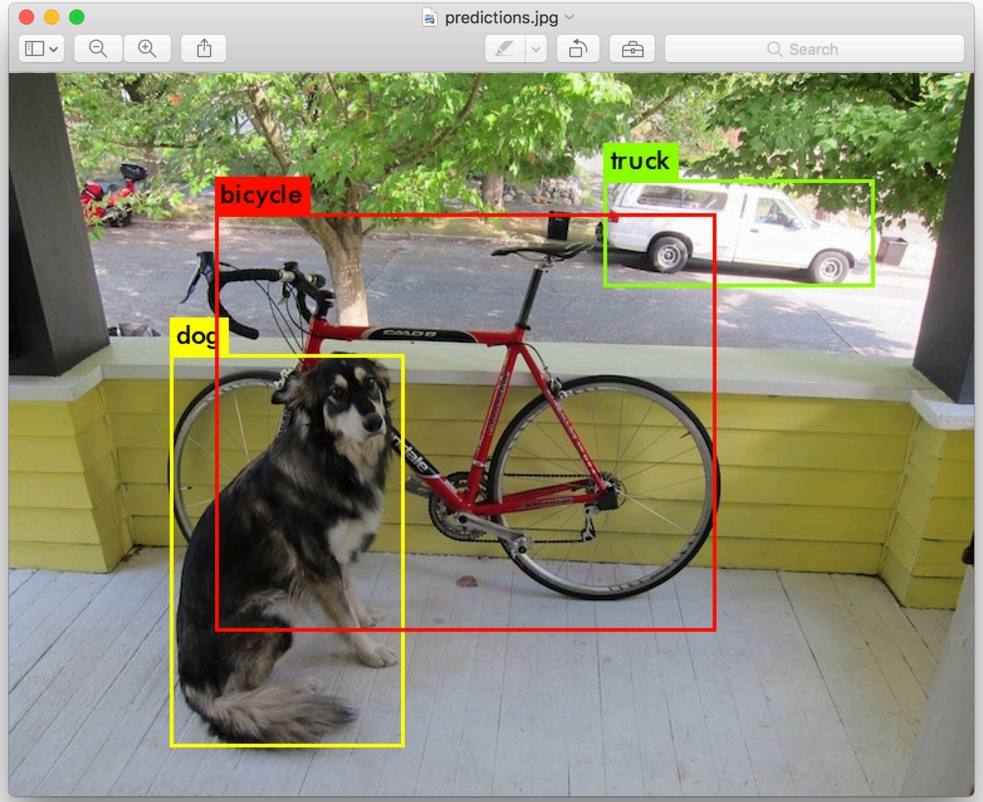
- Classification: object detection (YOLOv3 [YOLOv3])



Machine learning tasks

- Classification: object detection (YOLOv3 [YOLOv3])

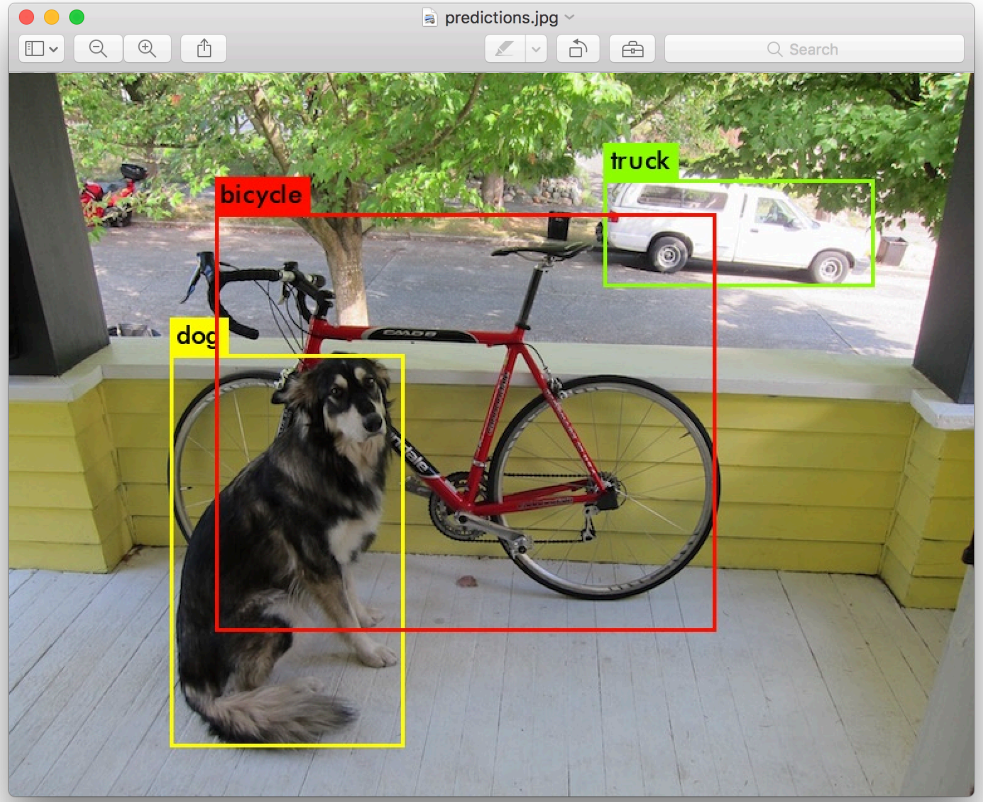
Task 1: Class recognition



Machine learning tasks

- Classification: object detection (YOLOv3 [YOLOv3])

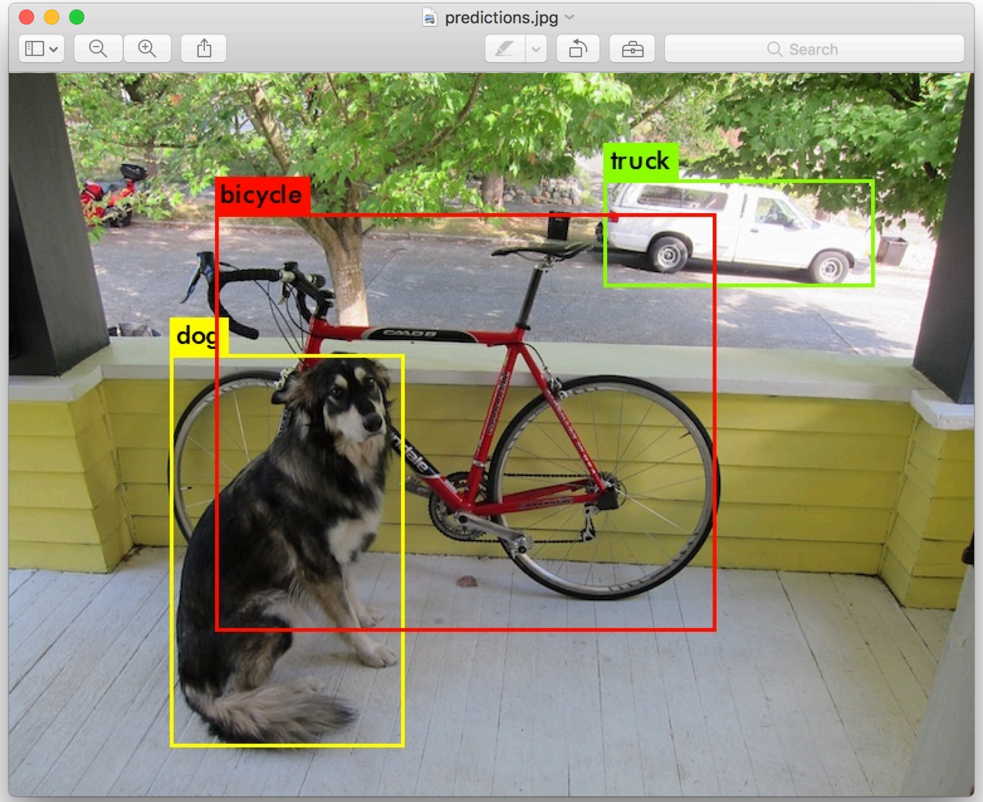
Task 1: Class recognition
Task 2: Localization



Machine learning tasks

- Classification: object detection (YOLOv3 [YOLOv3])

Task 1: Class recognition
Task 2: Localization

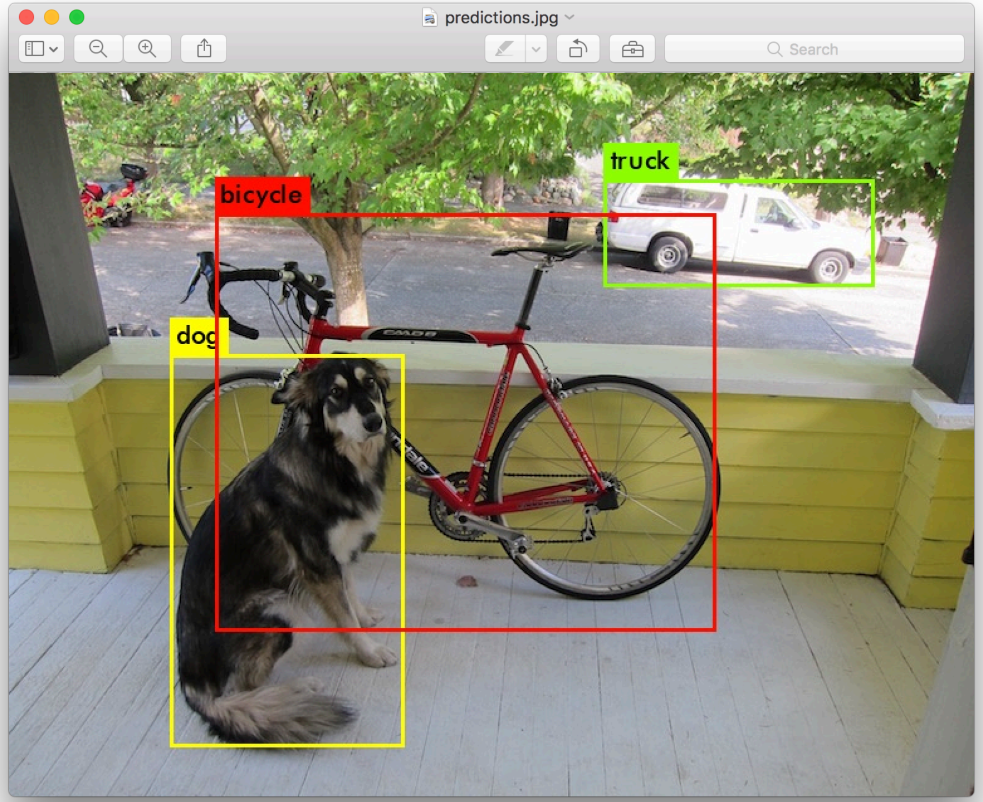


[A demo video]() of **YOLOv3** from https://pjreddie.com/darknet/yolo/

Machine learning tasks

- Classification: object detection (YOLOv3 [YOLOv3])

Task 1: Class recognition
Task 2: Localization



Q: How we can use it in practice?

A demo video of **YOLOv3** from https://pjreddie.com/darknet/yolo/

Machine learning tasks

- Classification: camera translate app

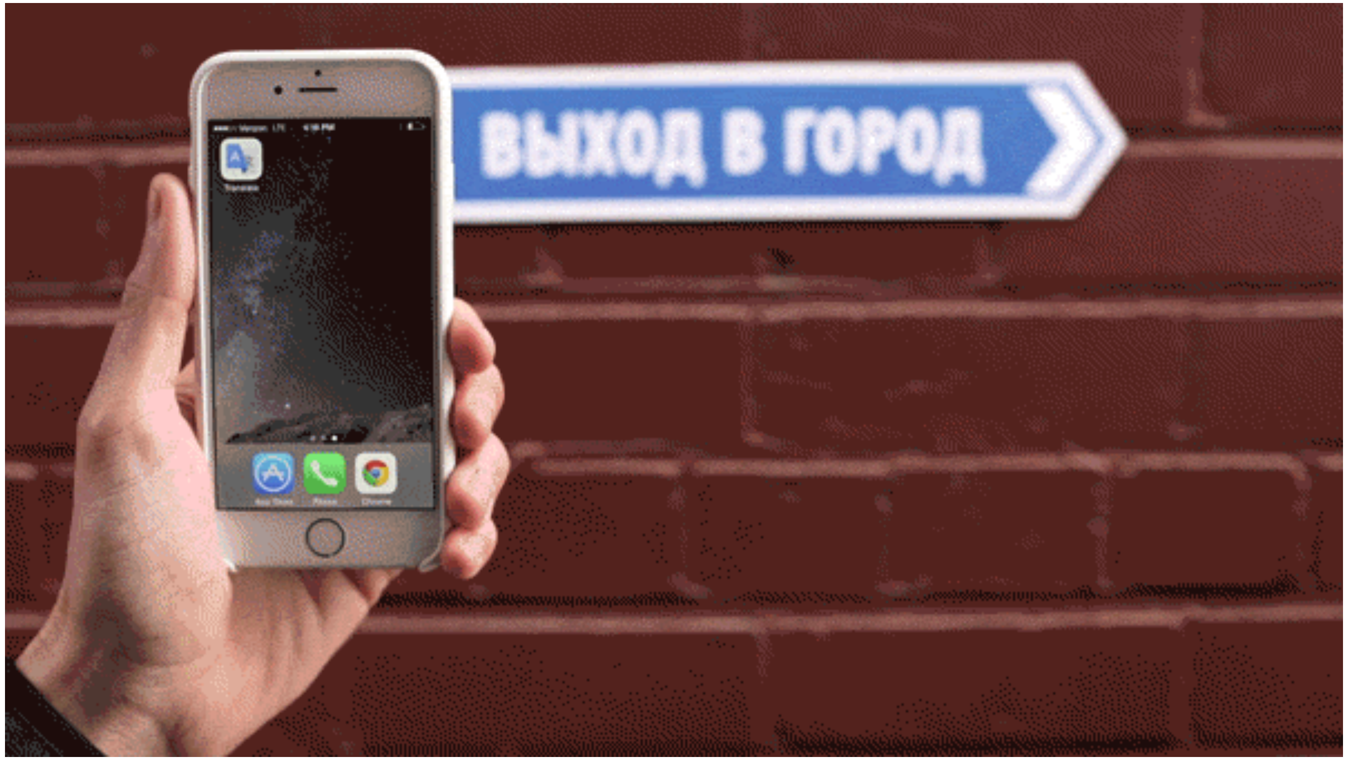


Image from https://petapixel.com/2015/01/14/googles-translate-app-can-now-use-camera-translate-world-real-time/

Machine learning tasks

- Classification: camera translate app (and other assistance using object detection?)

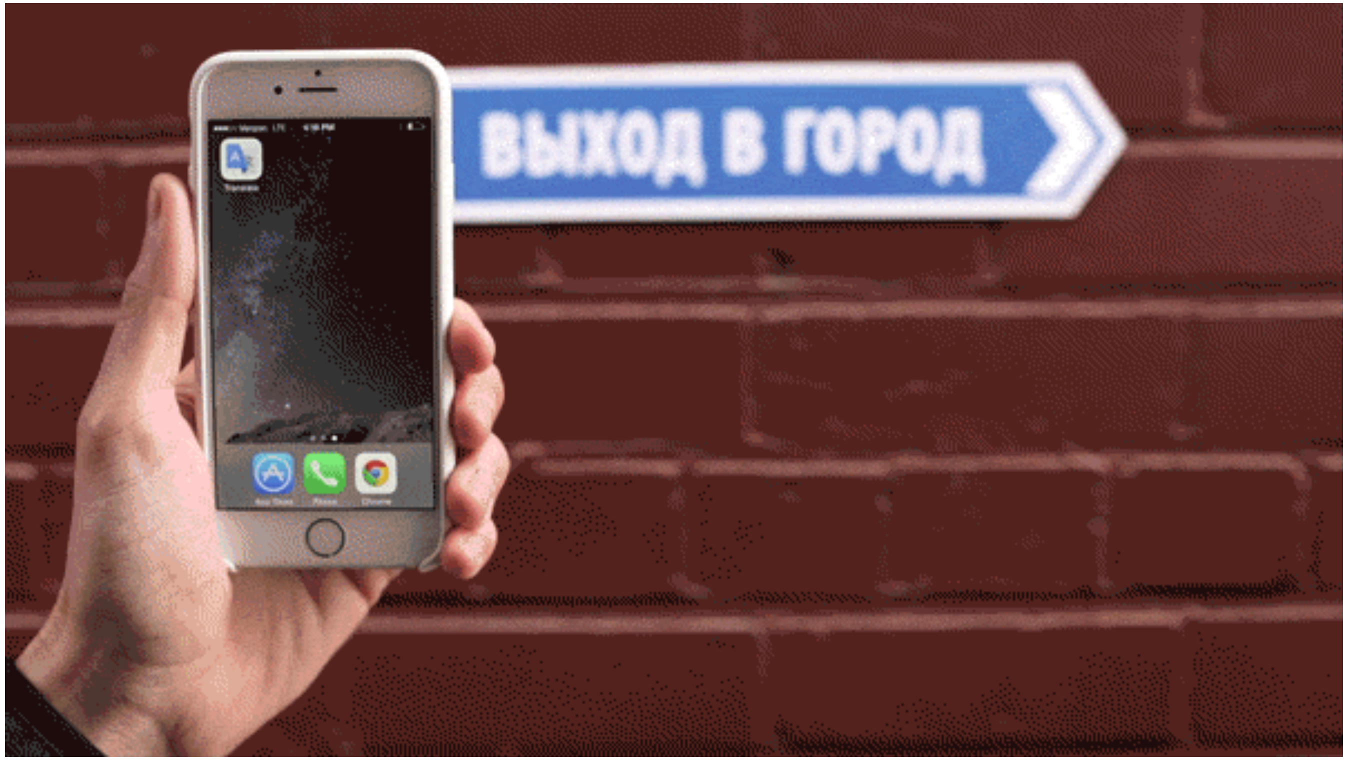


Image from https://petapixel.com/2015/01/14/googles-translate-app-can-now-use-camera-translate-world-real-time/

Machine learning tasks

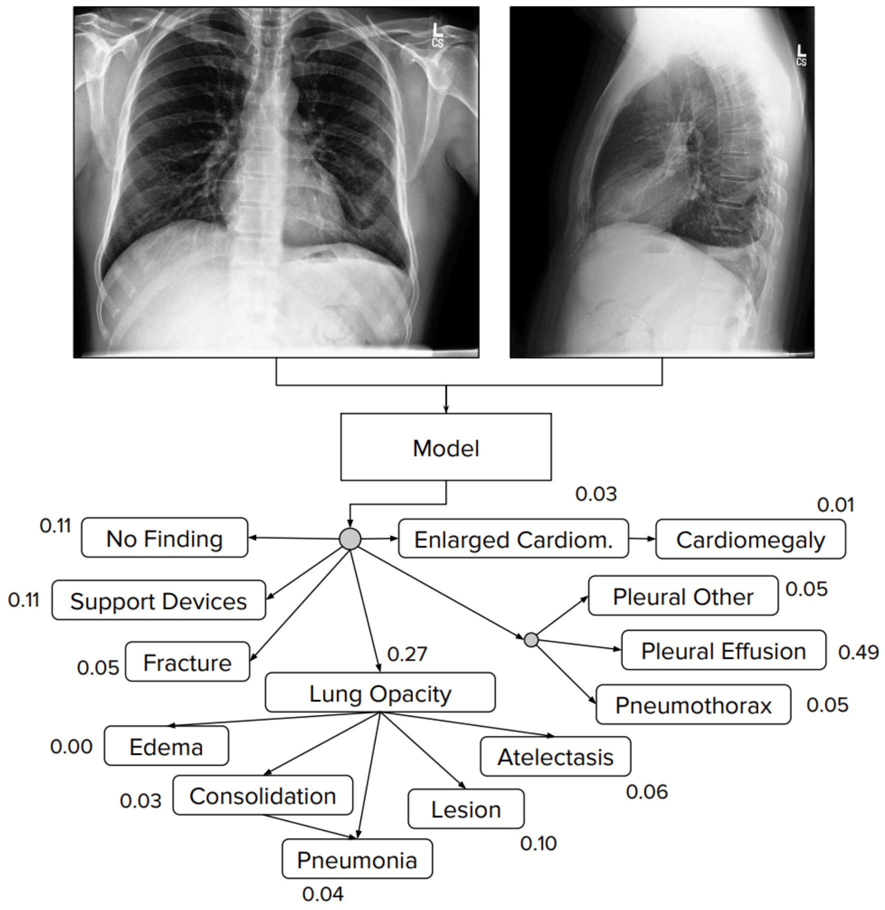
- Classification: camera translate app (and other assistance using object detection?)

Image from https://www.indianweb2.com/2014/07/two-indians-have-made-iron-man-jarvis.html

Machine learning tasks

- Classification: medical imaging

An example:
Automated chest radiograph interpretation.
Figure from [CheXpert].
CheXpert data [available](#).

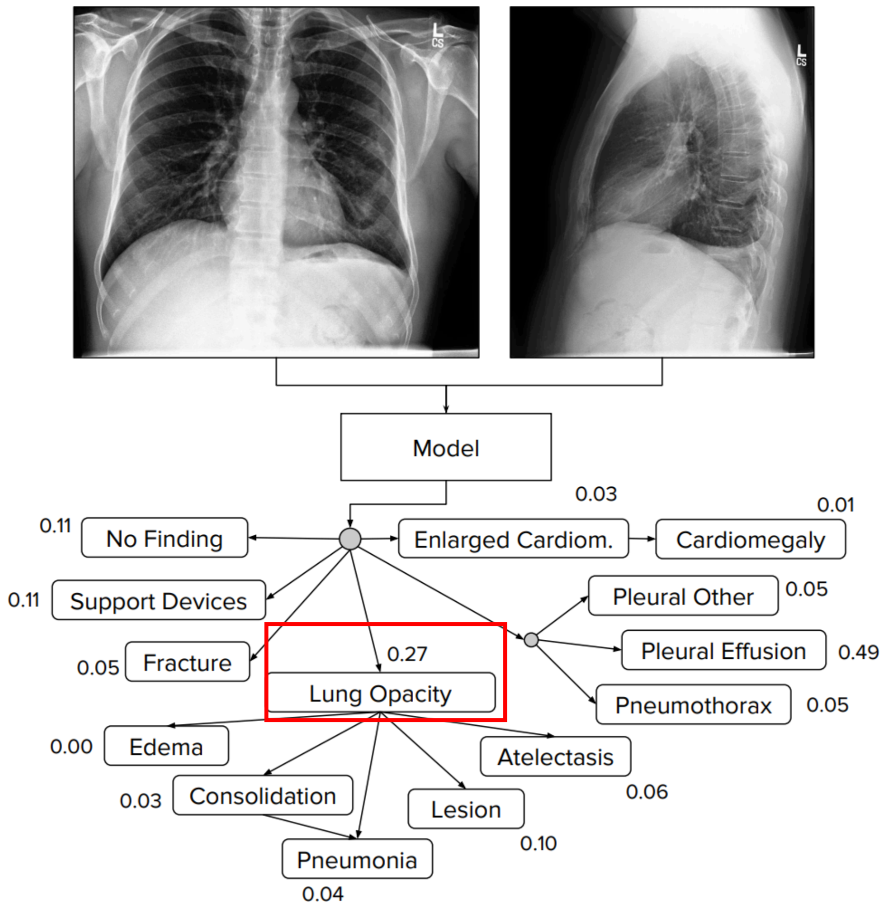


Machine learning tasks

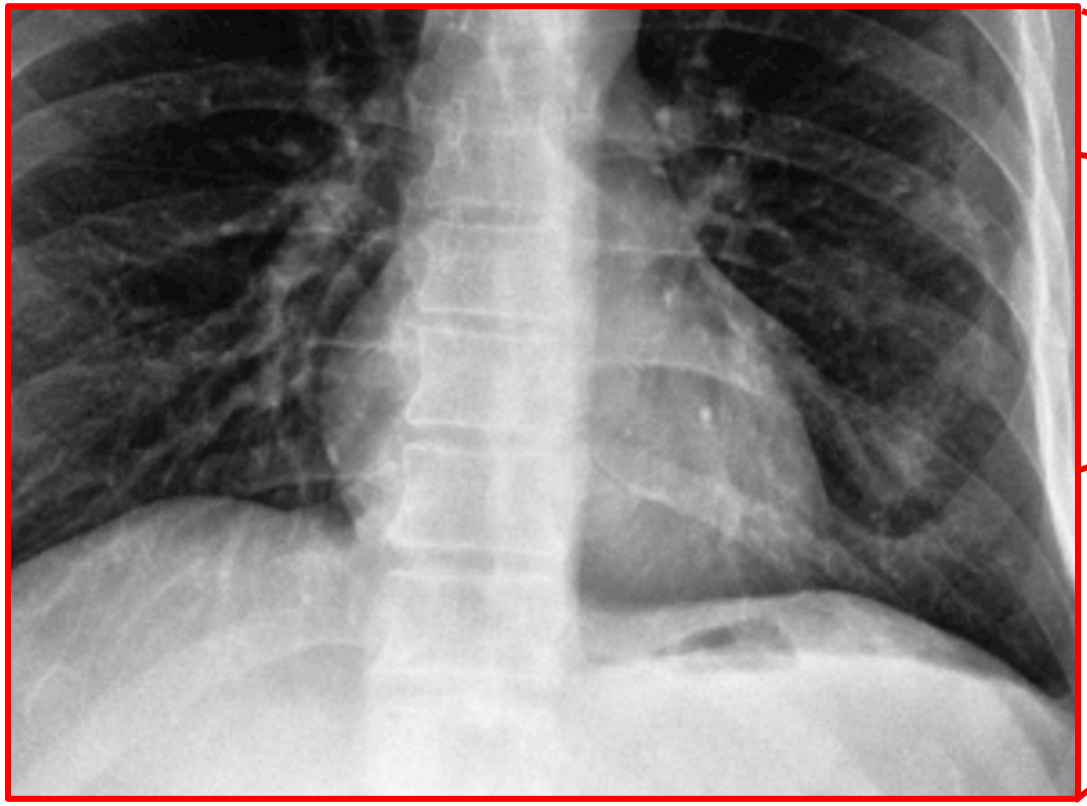
- Classification: medical imaging

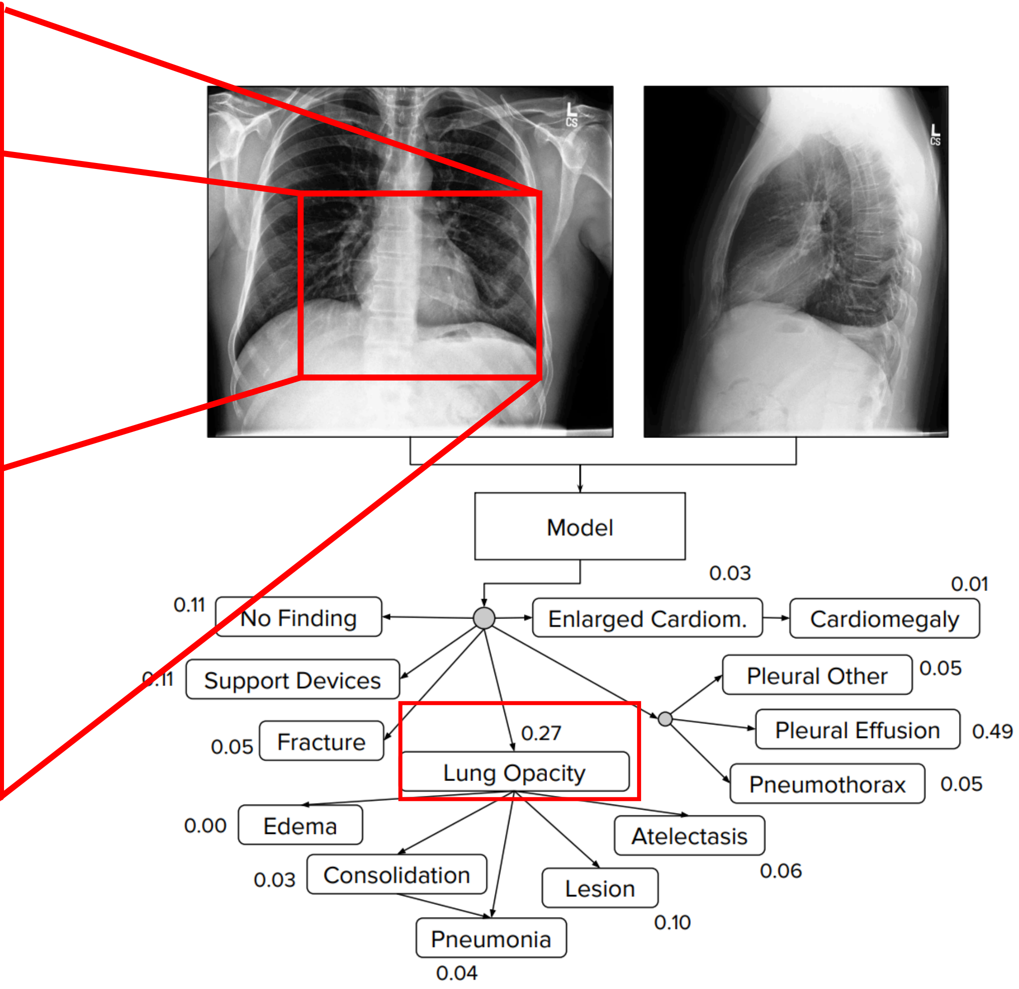
An example:
Automated chest radiograph interpretation.
Figure from [CheXpert].
CheXpert data available.

Q: why we need the automated interpretation?



Machine learning tasks

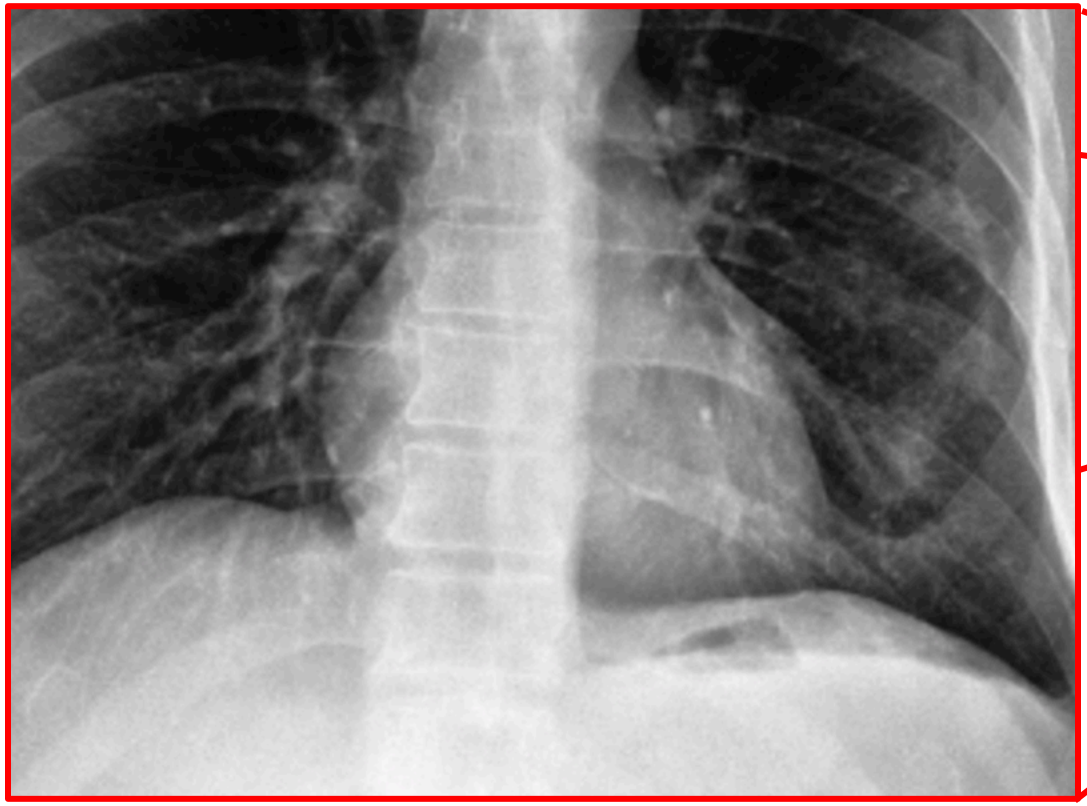


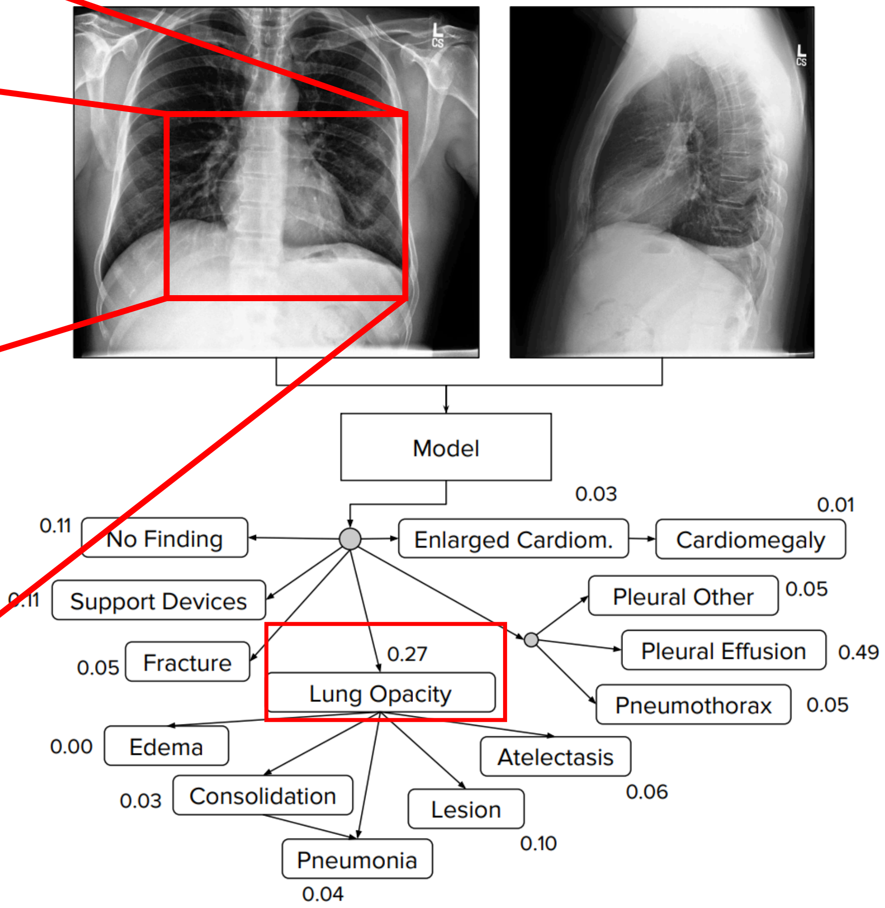


Please help to point out:

1. Whether there is lung opacity?
2. Why?

Machine learning tasks





Please help to point out:

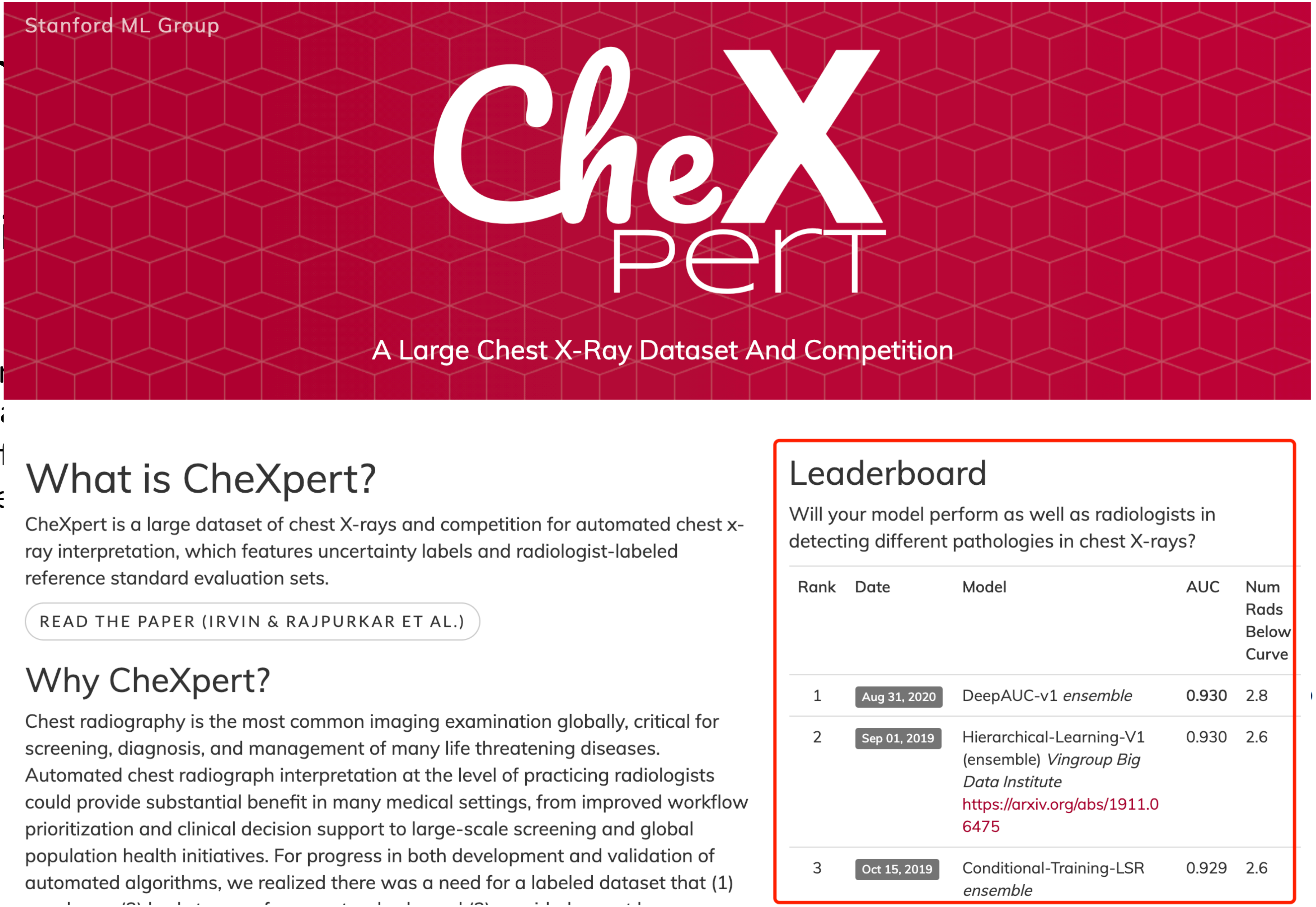
1. Whether there is lung opacity?
2. Why?

Difficult without expertise and sufficient time

Mach

- Classifi

An exa
Automa
Figure f
ChexPe



Stanford ML Group

CheXpert

A Large Chest X-Ray Dataset And Competition

What is CheXpert?

CheXpert is a large dataset of chest X-rays and competition for automated chest x-ray interpretation, which features uncertainty labels and radiologist-labeled reference standard evaluation sets.

READ THE PAPER (IRVIN & RAJPURKAR ET AL.)

Why CheXpert?

Chest radiography is the most common imaging examination globally, critical for screening, diagnosis, and management of many life threatening diseases. Automated chest radiograph interpretation at the level of practicing radiologists could provide substantial benefit in many medical settings, from improved workflow prioritization and clinical decision support to large-scale screening and global population health initiatives. For progress in both development and validation of automated algorithms, we realized there was a need for a labeled dataset that (1)

Leaderboard

Will your model perform as well as radiologists in detecting different pathologies in chest X-rays?

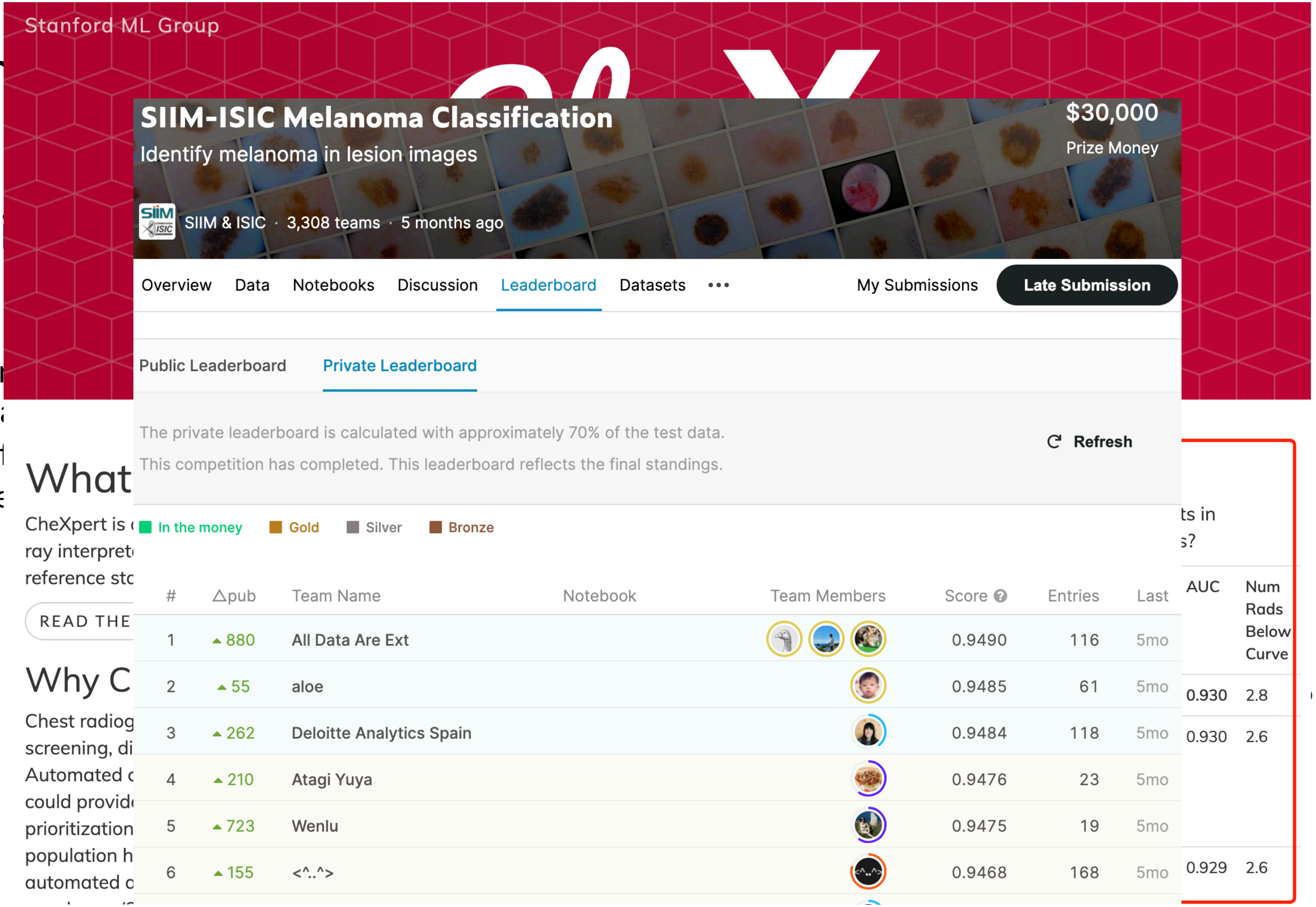
Rank	Date	Model	AUC	Num Rads Below Curve
1	Aug 31, 2020	DeepAUC-v1 *ensemble*	0.930	2.8
2	Sep 01, 2019	Hierarchical-Learning-V1 (ensemble) *Vingroup Big Data Institute* https://arxiv.org/abs/1911.06475	0.930	2.6
3	Oct 15, 2019	Conditional-Training-LSR *ensemble*	0.929	2.6

Some competitions

Mach

- Classifi

An exa
Automa
Figure f
CheXpe



#	Δpub	Team Name	Notebook	Team Members	Score	Entries	Last
1	▲ 880	All Data Are Ext			0.9490	116	5mo
2	▲ 55	aloe			0.9485	61	5mo
3	▲ 262	Deloitte Analytics Spain			0.9484	118	5mo
4	▲ 210	Atagi Yuya			0.9476	23	5mo
5	▲ 723	Wenlu			0.9475	19	5mo
6	▲ 155	<^..^>			0.9468	168	5mo

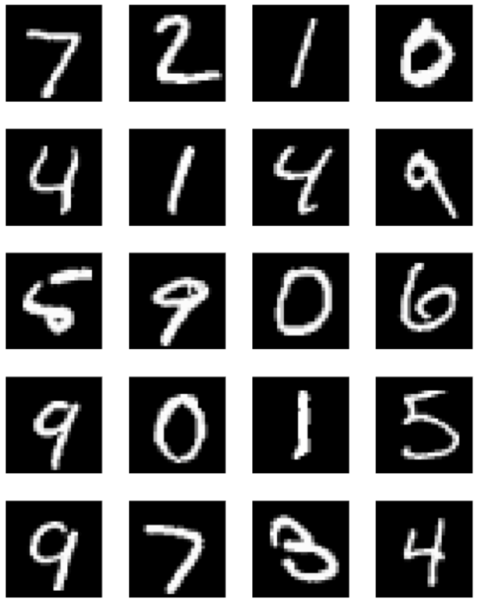
AUC	Num Rads Below Curve
0.930	2.8
0.930	2.6
0.929	2.6

Kaggle SIIM-ISIC competition: https://www.kaggle.com/c/siim-isic-melanoma-classification/discussion/163944

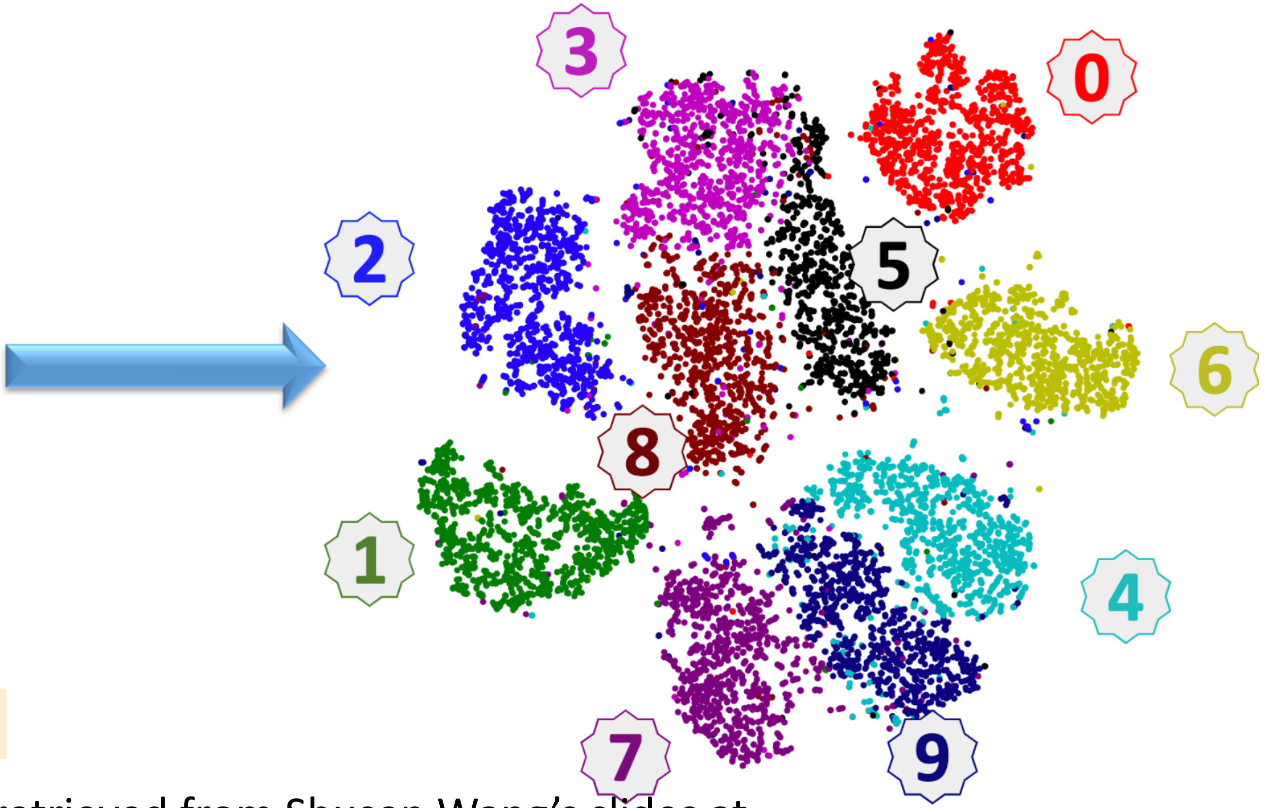
Machine learning tasks

- Clustering: unsupervised autoencoder

Raw image space
(grey scale matrices)



60K images (size 28×28)

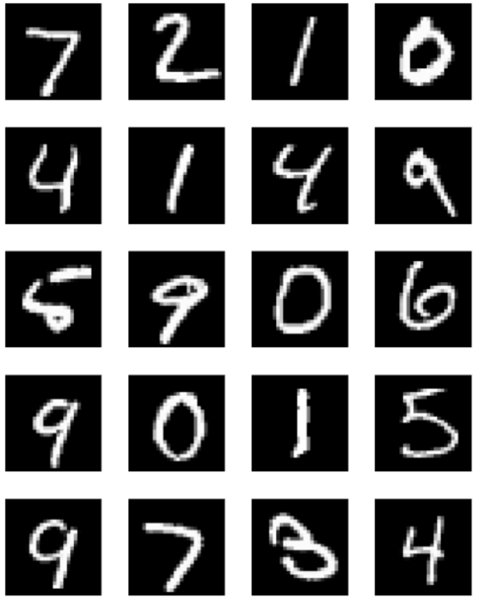


Autoencoder for hand-written digit data, retrieved from Shusen Wang's slides at
https://github.com/wangshusen/DeepLearning/blob/master/Slides/1_ML_Basics.pdf

Machine learning tasks

- Clustering: unsupervised autoencoder

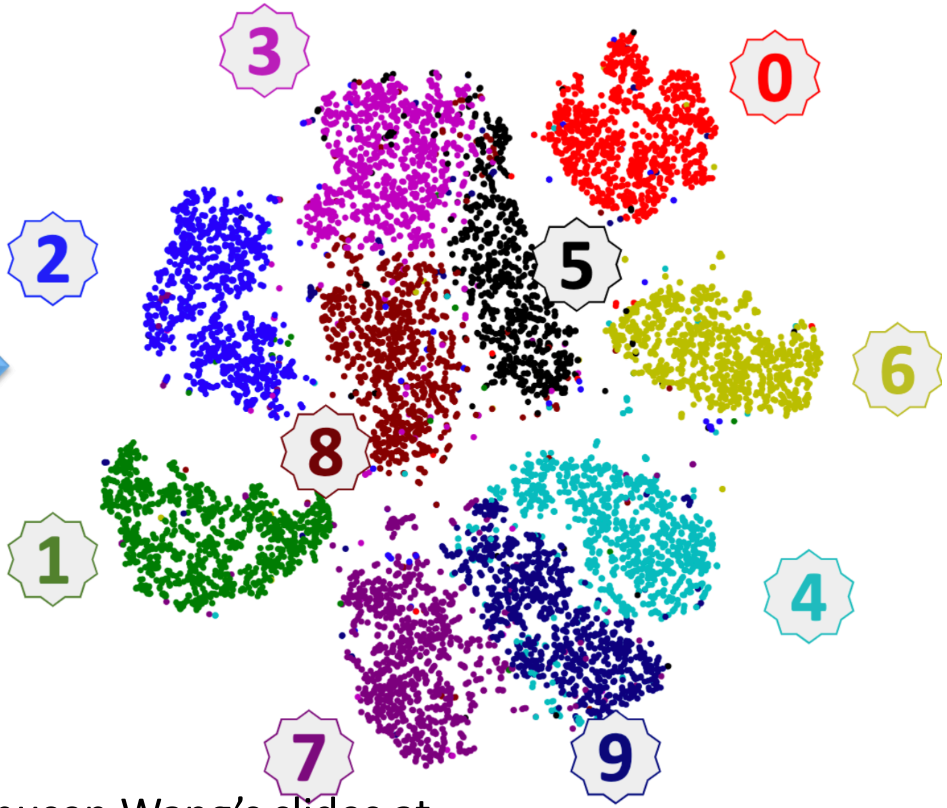
Raw image space
(grey scale matrices)



60K images (size 28×28)

Mapping to a new space

But why?

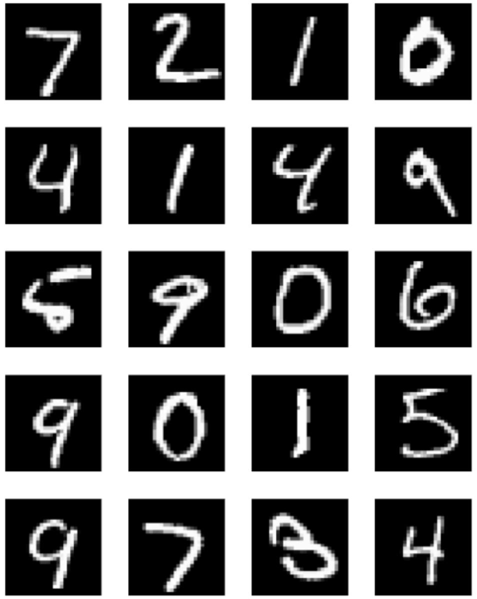


Autoencoder for hand-written digit data, retrieved from Shusen Wang's slides at https://github.com/wangshusen/DeepLearning/blob/master/Slides/1_ML_Basics.pdf

Machine learning tasks

- Clustering: unsupervised autoencoder

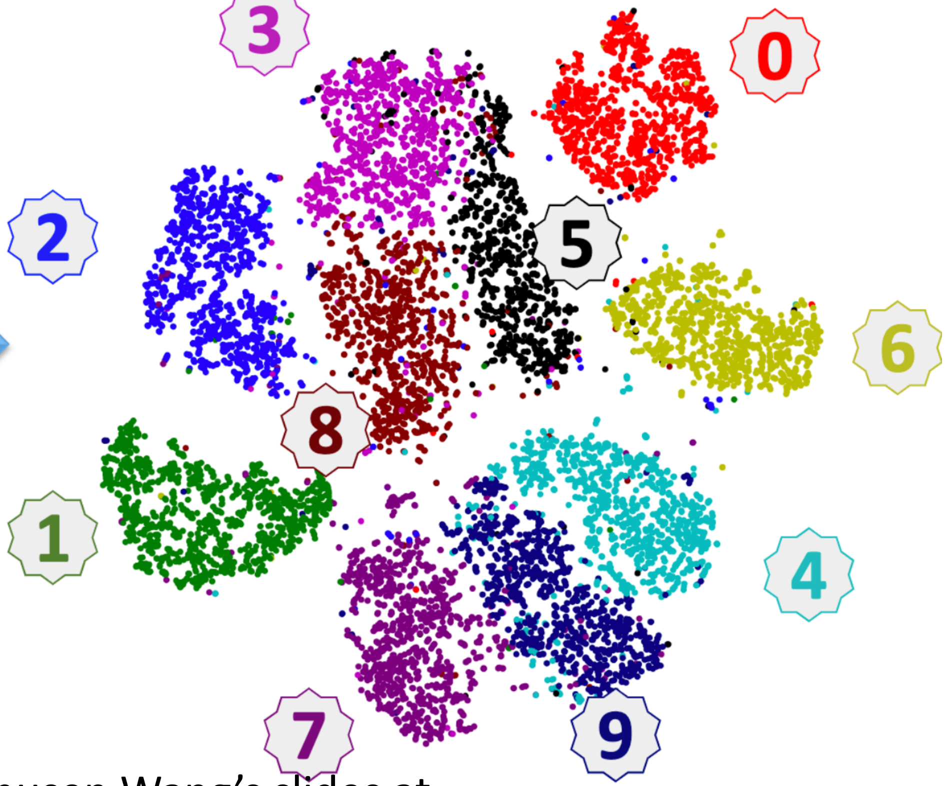
Raw image space
(grey scale matrices)



60K images (size 28×28)

Mapping to
a new space

X-y coordinates:
Easy for analysis,
e.g., classification



Autoencoder for hand-written digit data, retrieved from Shusen Wang's slides at
https://github.com/wangshusen/DeepLearning/blob/master/Slides/1_ML_Basics.pdf

Machine learning tasks

- Clustering: social network detection

From
1. Friendships on FB
2. Co-purchasing on Amazon

…

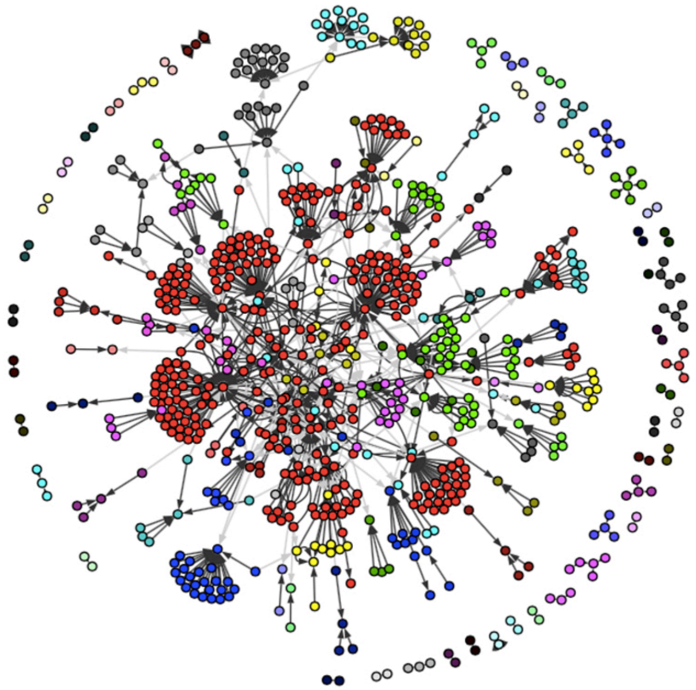


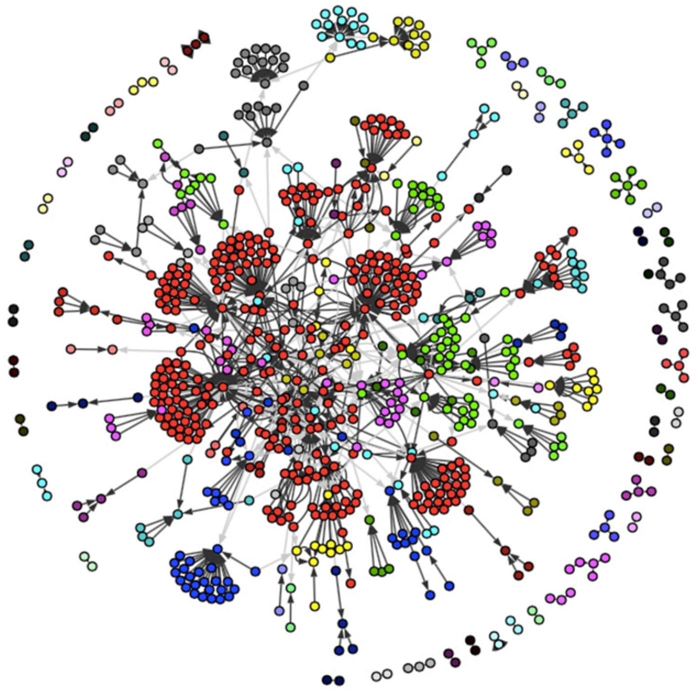
Image from https://www.kernix.com/article/community-detection-in-social-networks/

Machine learning tasks

- Clustering: social network detection

From
1. Friendships on FB
2. Co-purchasing on Amazon

…



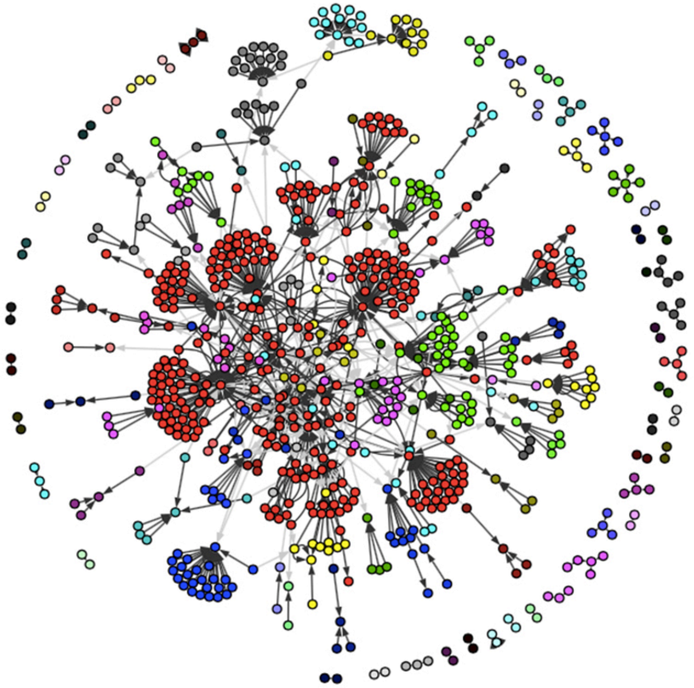
???

Image from https://www.kernix.com/article/community-detection-in-social-networks/

Machine learning tasks

- Clustering: social network detection

From
1. Friendships on FB
2. Co-purchasing on Amazon
…



Target
1. Recommendation systems
2. Identify groups in community health planning initiative
…

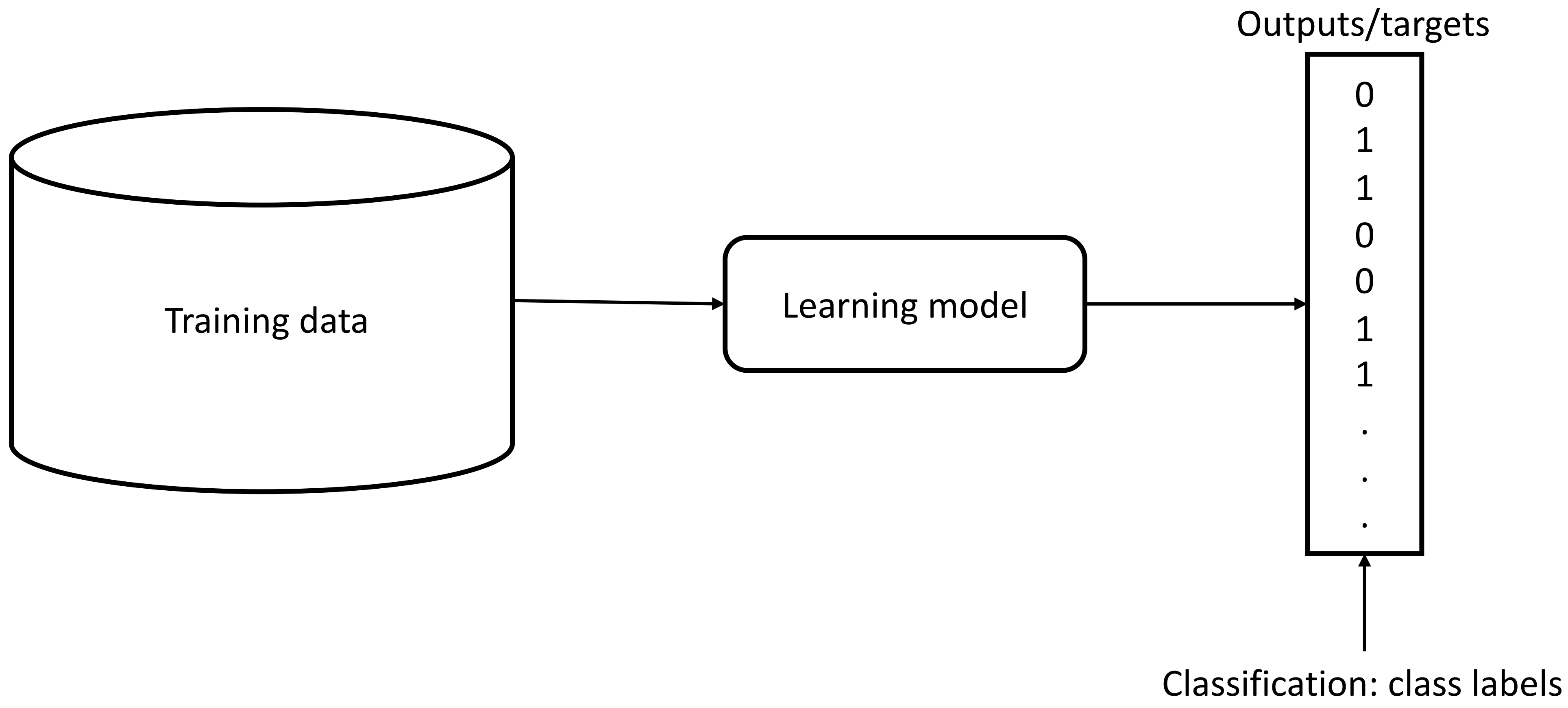
Image from https://www.kernix.com/article/community-detection-in-social-networks/

Machine learning tasks

- Regression: different setting from classification

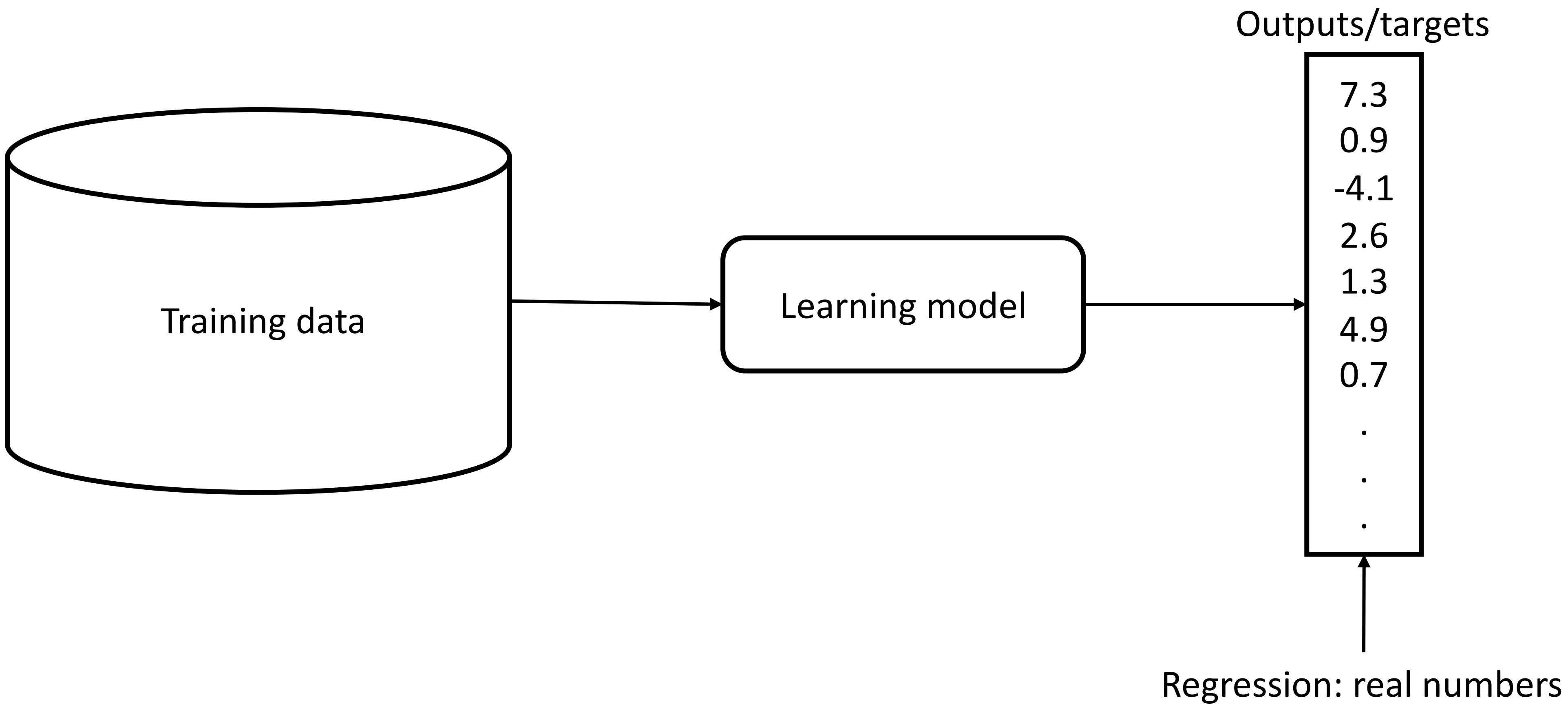
Machine learning tasks

- Regression: different setting from classification



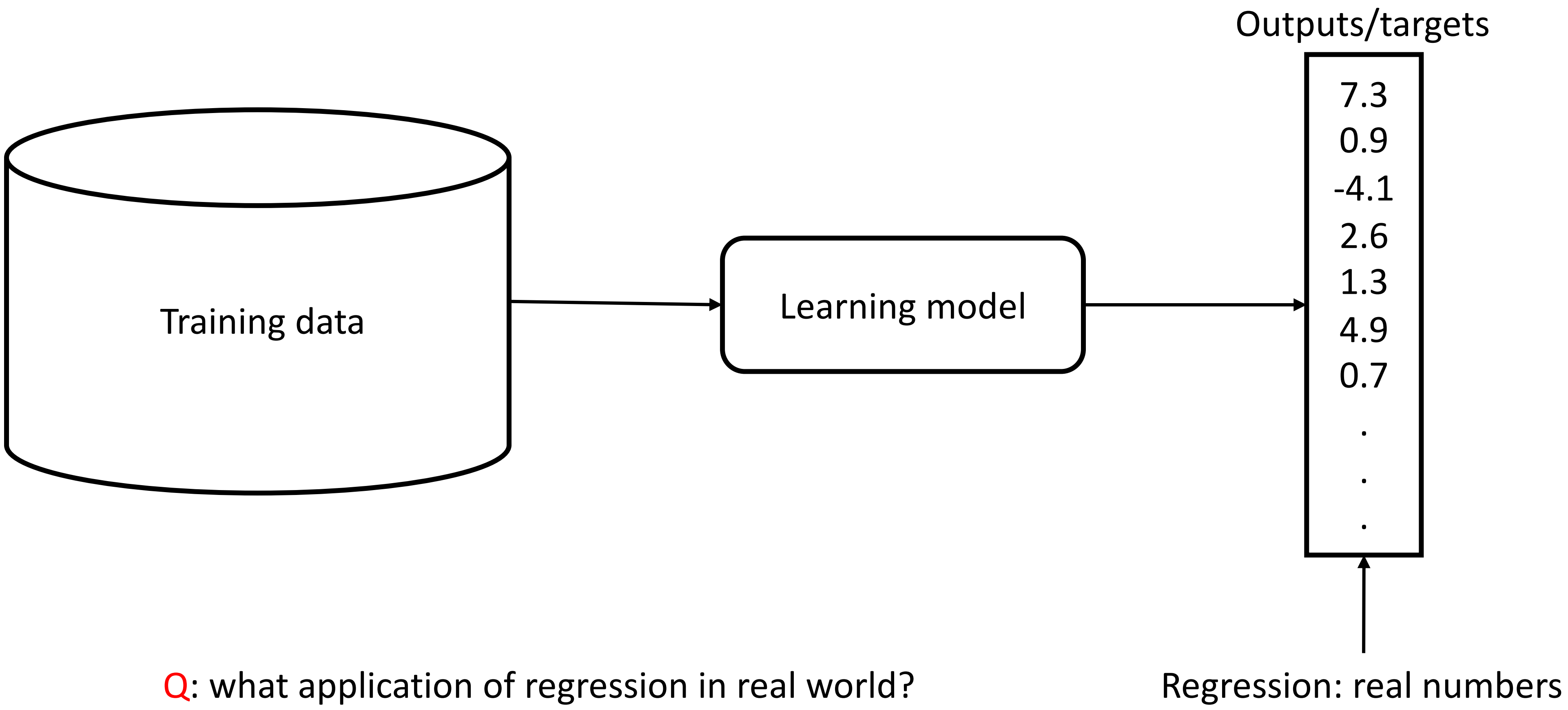
Machine learning tasks

- Regression: different setting from classification



Machine learning tasks

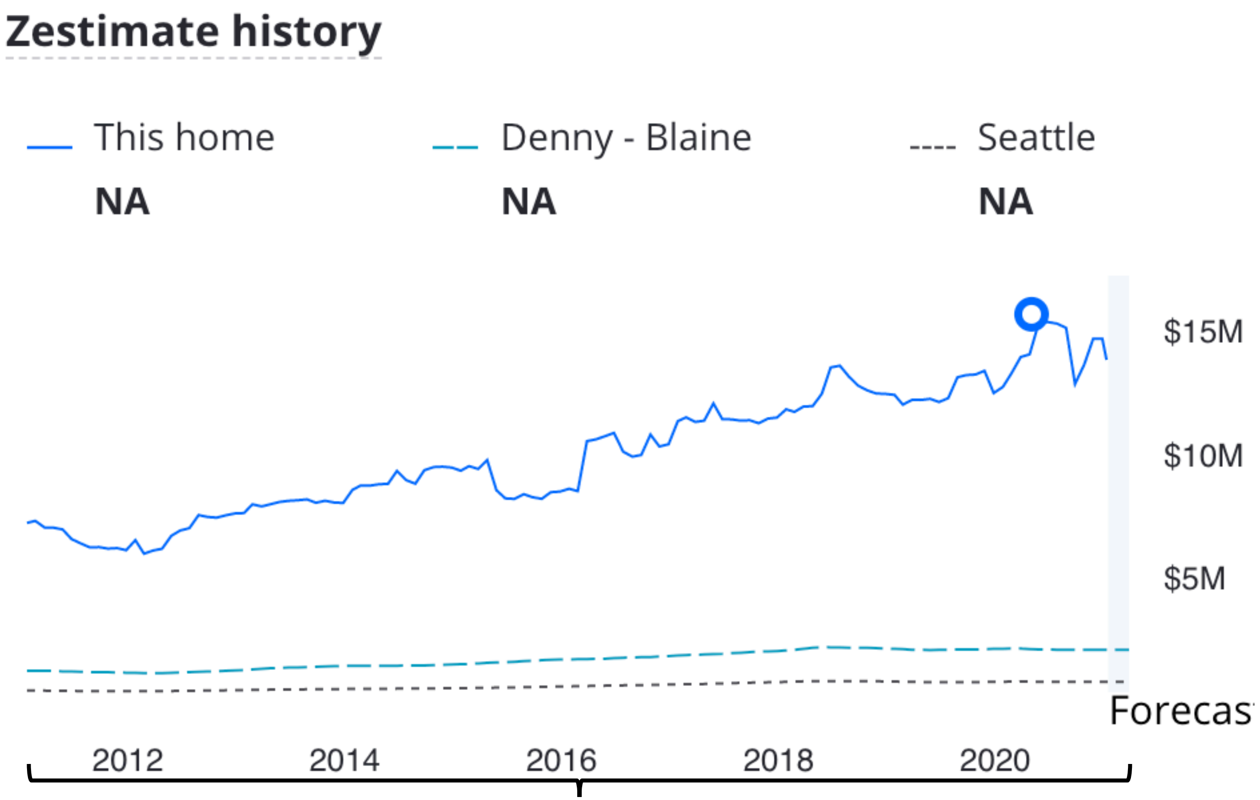
- Regression: different setting from classification



Q: what application of regression in real world?

Machine learning tasks

- Regression: house price estimation



No trade

Showcase how to determine a model (structure and parameters)

Components of machine learning paradigm

Training data

How to choose/generate useful features?

Learning model for bus recognition

How to determine this model?

Prediction: bus

Testing data

Pre-defined settings/tasks

From real applications: providing demand/requirement of problems

Components of machine learning paradigm

Training data

How to choose/generate useful features?

Learning model for bus recognition

1. Determine model structure
2. Determine model parameters

Prediction: bus

Testing data

Pre-defined settings/tasks

From real applications: providing demand/requirement of problems

House price model structure: linear model

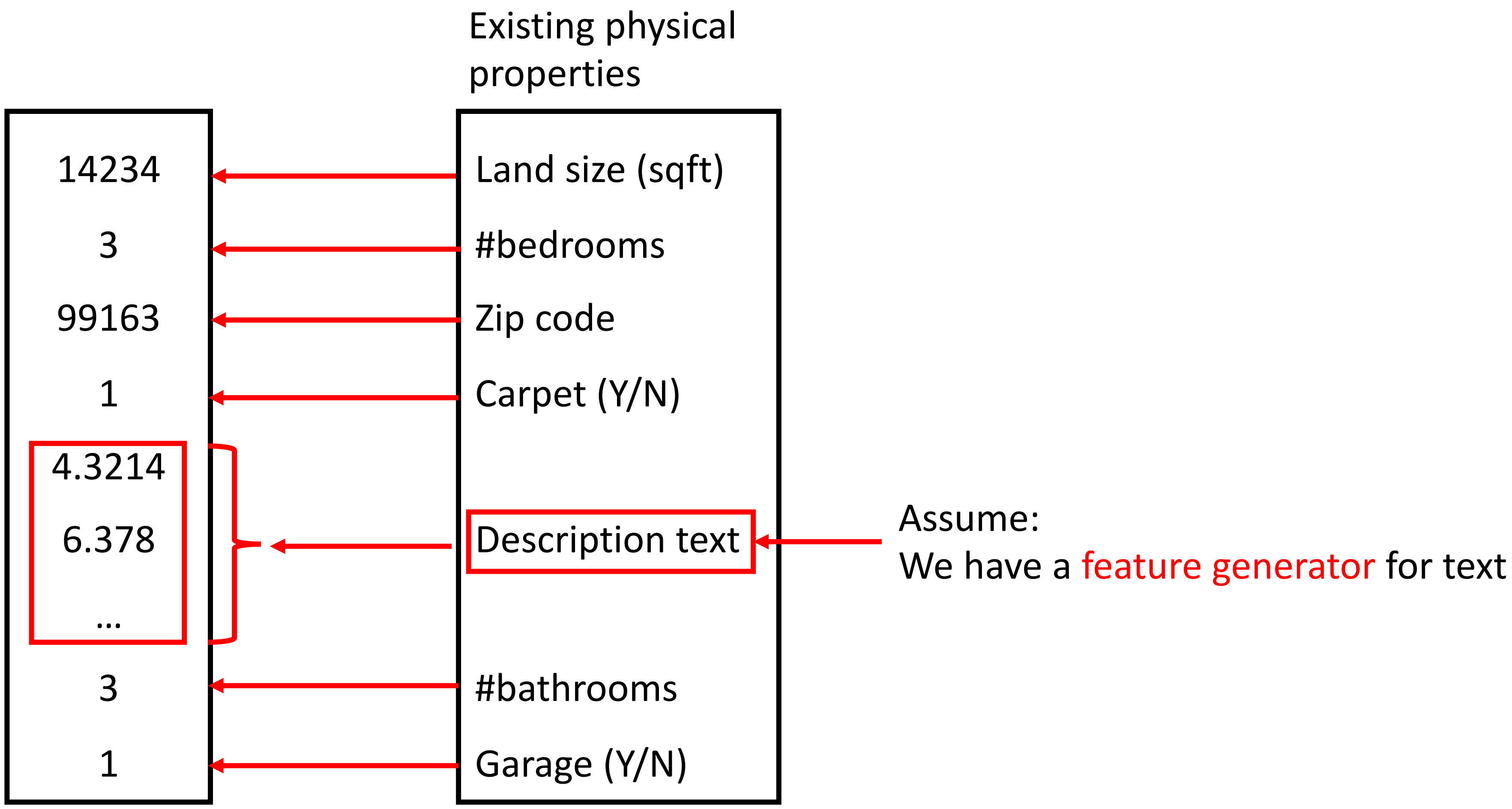
- Features:
  - Home characteristics: lot size, location, #bedrooms
  - Unique features: hardwood floors, granite countertops or a landscaped backyard
  - On-market data: listing price, description, days on the market
  - Off-market data: tax assessments, prior sales

House price model structure: linear model

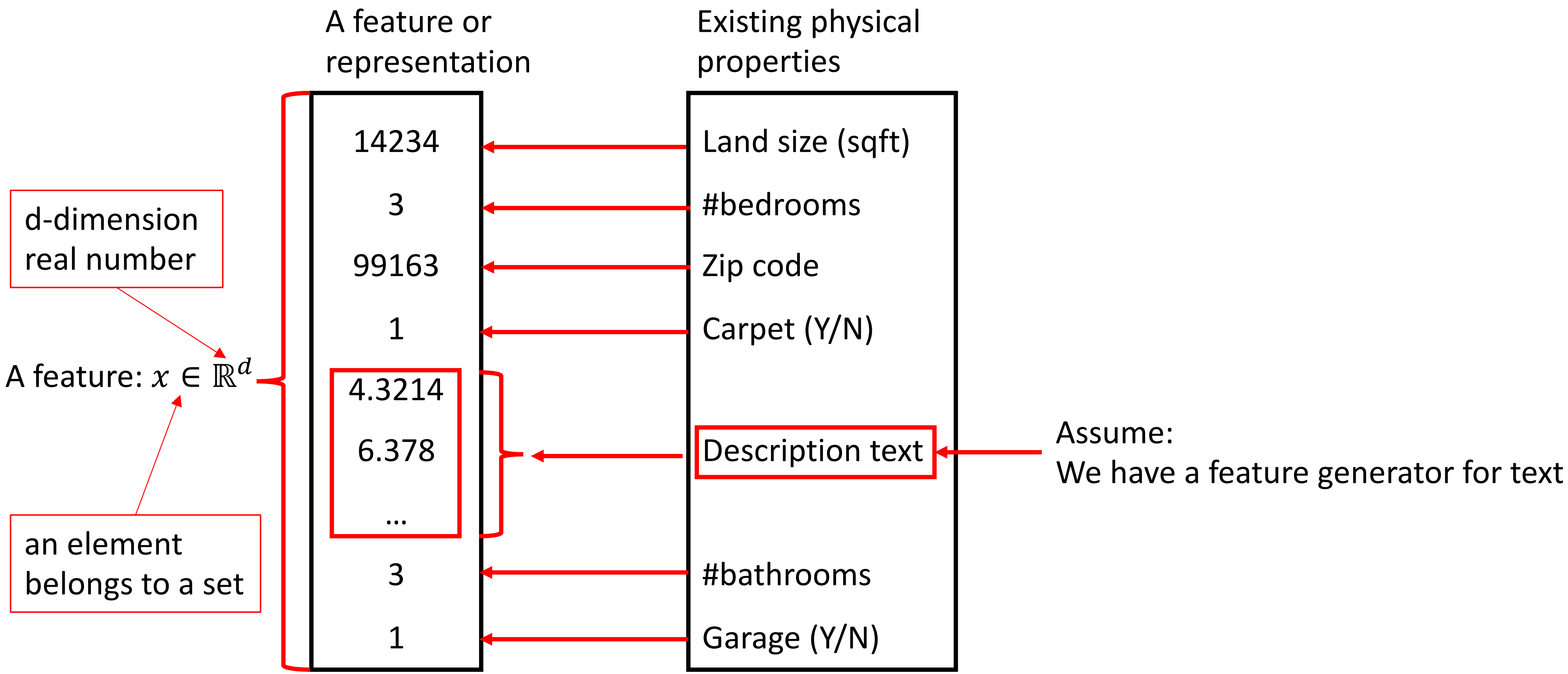
Existing physical properties

Land size (sqft)
#bedrooms
Zip code
Carpet (Y/N)
Description text
#bathrooms
Garage (Y/N)

House price model structure: linear model



House price model structure: linear model



House price model structure: linear model

- Prediction model: a linear function

$$f(w; x_i) = \underbrace{x_i{}'w}_{\text{Weight}} + \underbrace{b}_{\text{Bias}}$$

House price model structure: linear model

- Prediction model: a linear function

$$\underbrace{f(w; x_i) = x_i{}'w + b}_{} \xrightarrow{\text{approximate}} y_i\text{: given trading price}$$

Weight   Bias

House price model structure: linear model

- Prediction model: a linear function

$$\underbrace{f(w; x_i) = x_i{}'w + b}_{\text{Weight, Bias}} \xrightarrow{\text{approximate}} y_i\text{: given trading price}$$

Weight Bias

Vector (n-dim) $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$

Matrix ($n \times d$) $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1d} \\ a_{21} & a_{22} & \cdots & a_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nd} \end{bmatrix}$

Row and columns $\mathbf{A} = \begin{bmatrix} \mathbf{a}_{:1} & \mathbf{a}_{:2} & \cdots & \mathbf{a}_{:d} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_{1:} \\ \mathbf{a}_{2:} \\ \vdots \\ \mathbf{a}_{n:} \end{bmatrix}$

Transpose:
$x' = x^T$ $\begin{bmatrix} 6 & 4 & 24 \\ 1 & -9 & 8 \end{bmatrix}^T = \begin{bmatrix} 6 & 1 \\ 4 & -9 \\ 24 & 8 \end{bmatrix}$

House price model structure: linear model

- Prediction model: a linear function

How to measure approximation?

$$\underbrace{f(w; x_i) = x_i' \overset{\text{Weight}}{w} + \overset{\text{Bias}}{b}}\xrightarrow{\text{approximate}} y_i\text{: given trading price}$$

Weight

Bias

Vector (n-dim) $\quad \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$

Matrix ($n \times d$) $\quad \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1d} \\ a_{21} & a_{22} & \cdots & a_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nd} \end{bmatrix}$

Row and columns $\quad \mathbf{A} = \begin{bmatrix} \mathbf{a}_{:1} & \mathbf{a}_{:2} & \cdots & \mathbf{a}_{:d} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_{1:} \\ \mathbf{a}_{2:} \\ \vdots \\ \mathbf{a}_{n:} \end{bmatrix}$

Transpose:
$x' = x^T$ $\quad \begin{bmatrix} 6 & 4 & 24 \\ 1 & -9 & 8 \end{bmatrix}^T = \begin{bmatrix} 6 & 1 \\ 4 & -9 \\ 24 & 8 \end{bmatrix}$

House price model structure: linear model

- Prediction model: a linear function

How to measure approximation?

$$f(w; x_i) = x_i{}'w + b \xrightarrow{\text{approximate}} y_i\text{: given trading price}$$

Weight Bias

- Loss function

$$l(f(w; x_i) - y_i)$$

  - Square loss

$$l(a, b) = \frac{1}{2}(a - b)^2$$

  - Absolute loss

$$l(a, b) = |a\text{-}b|$$

House price model structure: linear model

- Prediction model: a linear function

How to measure approximation?

$$f(w; x_i) = x_i'w + b \xrightarrow{\text{approximate}} y_i\text{: given trading price}$$

Weight Bias

parameters

- Loss function

$$l(f(w; x_i) - y_i)$$

  - Square loss

$$l(a, b) = \frac{1}{2}(a - b)^2$$

  - Absolute loss

$$l(a, b) = |a\text{-}b|$$

A trick for merging bias:

$$x \leftarrow \begin{bmatrix} \mathbf{x}_i \\ 1 \end{bmatrix} \qquad w \leftarrow \begin{bmatrix} \mathbf{w} \\ b \end{bmatrix}$$

Determining model parameters

- An optimization problem (on training set)

n training data

$$min_w \frac{1}{n}\sum_{i=1}^{n} l(f(w;x_i),y_i) = \frac{1}{2n}\sum_{i=1}^{n}(x_i{}'w - y_i)^2$$

Determining model parameters

- An optimization problem (on training set)

n training data

$$min_w \frac{1}{n}\sum_{i=1}^{n} l(f(w;x_i),y_i) = \frac{1}{2n}\sum_{i=1}^{n}(x_i{}'w - y_i)^2$$

- 1-dimensional 1-data showcase with square loss

$$min_w \tfrac{1}{2}(wx - y)^2$$

- Analytical solution?

Determining model parameters

- An optimization problem (on training set)

n training data

$$min_w \frac{1}{n}\sum_{i=1}^{n} l(f(w;x_i),y_i) = \frac{1}{2n}\sum_{i=1}^{n}(x_i{}'w - y_i)^2$$

- 1-dimensional 1-data showcase with square loss

$$min_w ½(wx - y)^2$$

- Analytical solution?

$$w^* = y/x$$

(first order optimality)

Determining model parameters

- An optimization problem (on training set)

n training data

$$min_w \frac{1}{n}\sum_{i=1}^{n} l(f(w;x_i),y_i) = \frac{1}{2n}\sum_{i=1}^{n}(x_i{}'w - y_i)^2$$

- 1-dimensional multi-data square loss

$$min_w \frac{1}{2n}\sum_{i=1}^{n}(wx - y)^2$$

- Analytical solution?

$$w = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

Determining model parameters

- An optimization problem (on training set)

n training data

$$min_w \frac{1}{n}\sum_{i=1}^{n} l(f(w; x_i), y_i) = \frac{1}{2n}\sum_{i=1}^{n}(x_i{}'w - y_i)^2$$

- multi-dimensional multi-data square loss

$$min_w \frac{1}{2n}\sum_{i=1}^{n}(x_i{}'w - y_i)^2$$

- Analytical solution?

References

- [YOLOv3] Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." *arXiv preprint arXiv:1804.02767* (2018).
- [CheXpert] Irvin, Jeremy, Pranav Rajpurkar, Michael Ko, Yifan Yu, Silviana Ciurea-Ilcus, Chris Chute, Henrik Marklund et al. "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison." In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 590-597. 2019.