Intro to ML concepts

Aarti Singh

Machine Learning 10-315
Sept 2, 2020

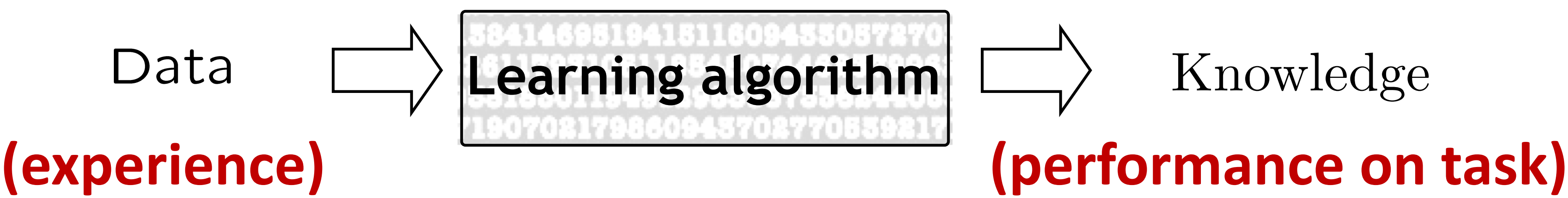
Logistical update

- Canvas fixed
  - Zoom links for lecture/recitation and office hours available on Canvas
  - Recording of lectures and recitations available at Zoom tab on Canvas
  - Piazza login directly
- Recitation on Friday Sept 4 – Probability distributions + optimization review and hands-on exercises
- QnA1 to be released TODAY

What is Machine Learning?

Design and Analysis of algorithms that

- improve their performance
- at some task
- with experience



Tasks, Experience, Performance

Machine Learning Tasks

Broad categories -

- **Supervised learning**

  Classification, Regression

- **Unsupervised learning**

  Density estimation, Clustering, Dimensionality reduction

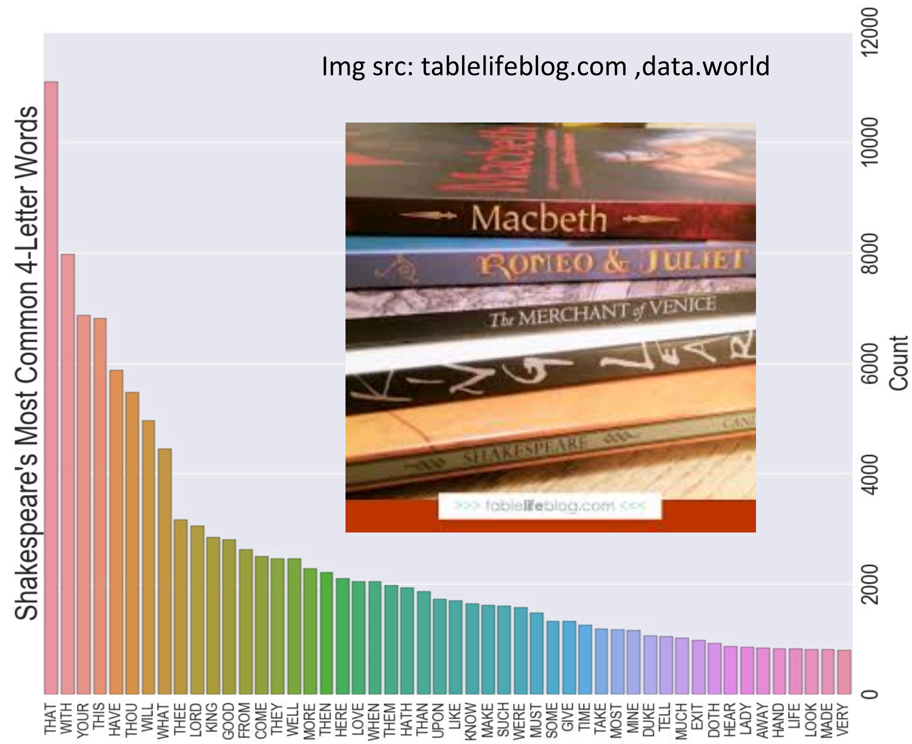
  Distribution

- Semi-supervised learning
- Active learning
- Reinforcement learning
- Many more …

Unsupervised Learning

Learning a Distribution

Bias of a coin

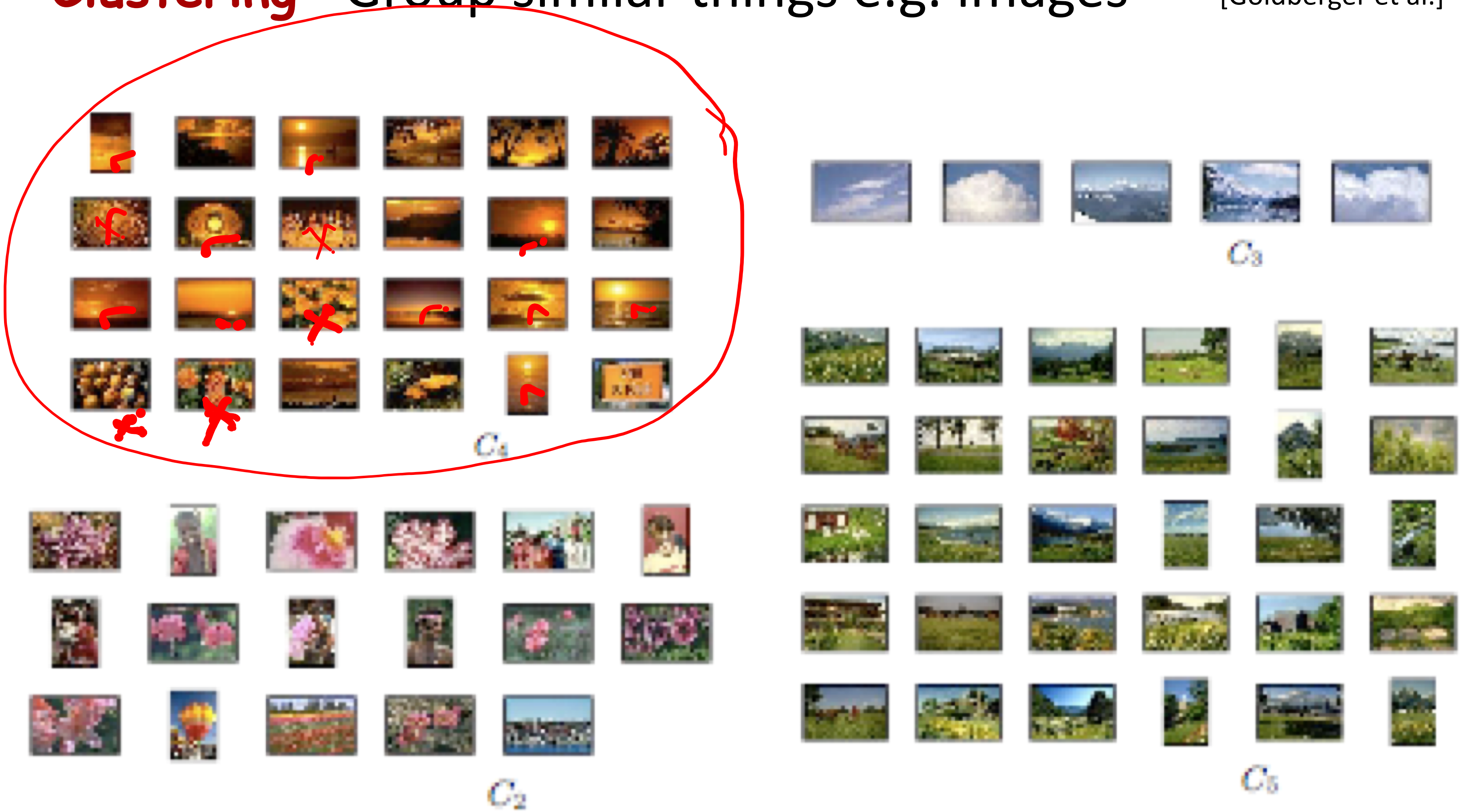


Distribution of words in text

- What other distribution would be interesting to learn?

Unsupervised Learning

Clustering - Group similar things e.g. images [Goldberger et al.]



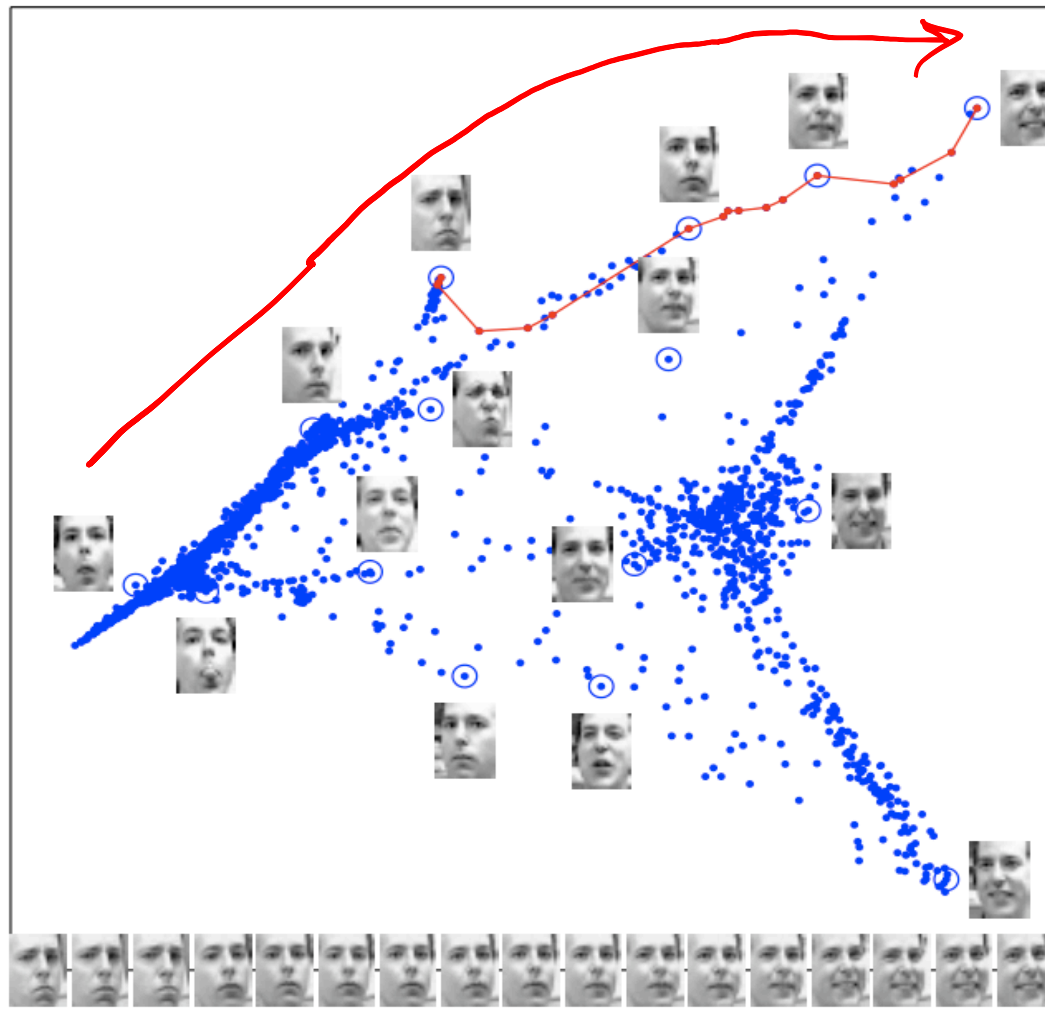
Unsupervised Learning

Dimensionality Reduction/Embedding

[Saul & Roweis '03]

Images have thousands or millions of pixels.

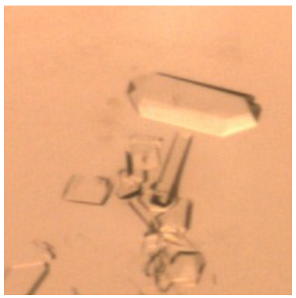
Can we give each image a small set of coordinates, such that similar images are near each other?



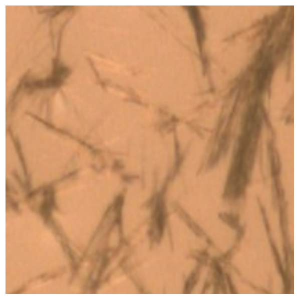
Tasks, **Experience**, Performance

Experience = Training Data

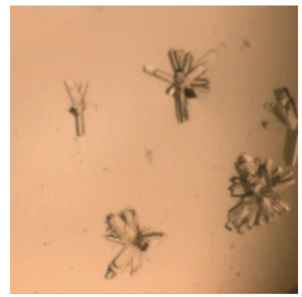
Task: Learning stage of protein crystallization



Crystal

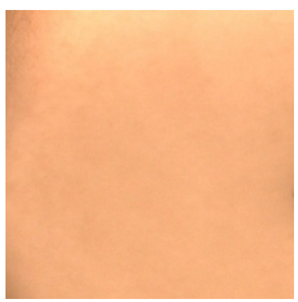


Needle

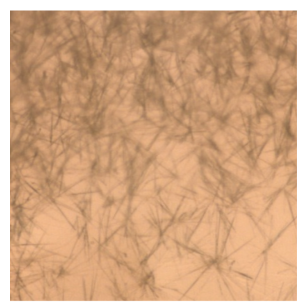


Tree

Tree

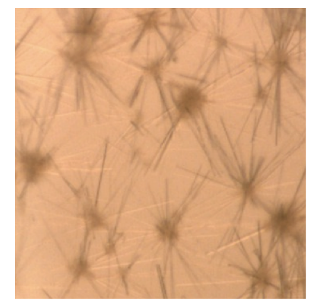


Empty



Needle

Experience

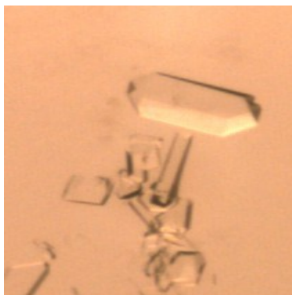


?

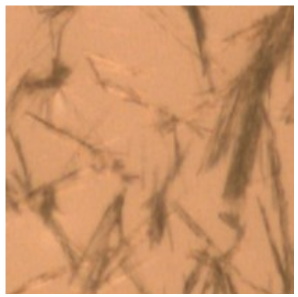
Performance

Training Data vs. Test Data

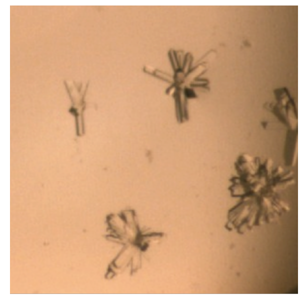
Task: Learning stage of protein crystallization



Crystal

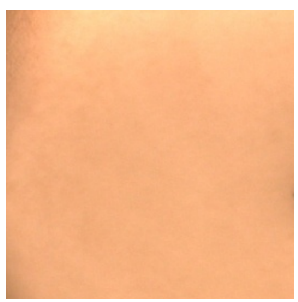


Needle

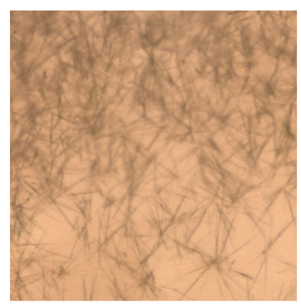


Tree

Tree

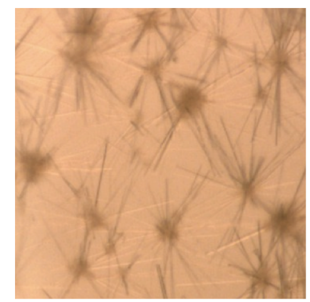


Empty



Needle

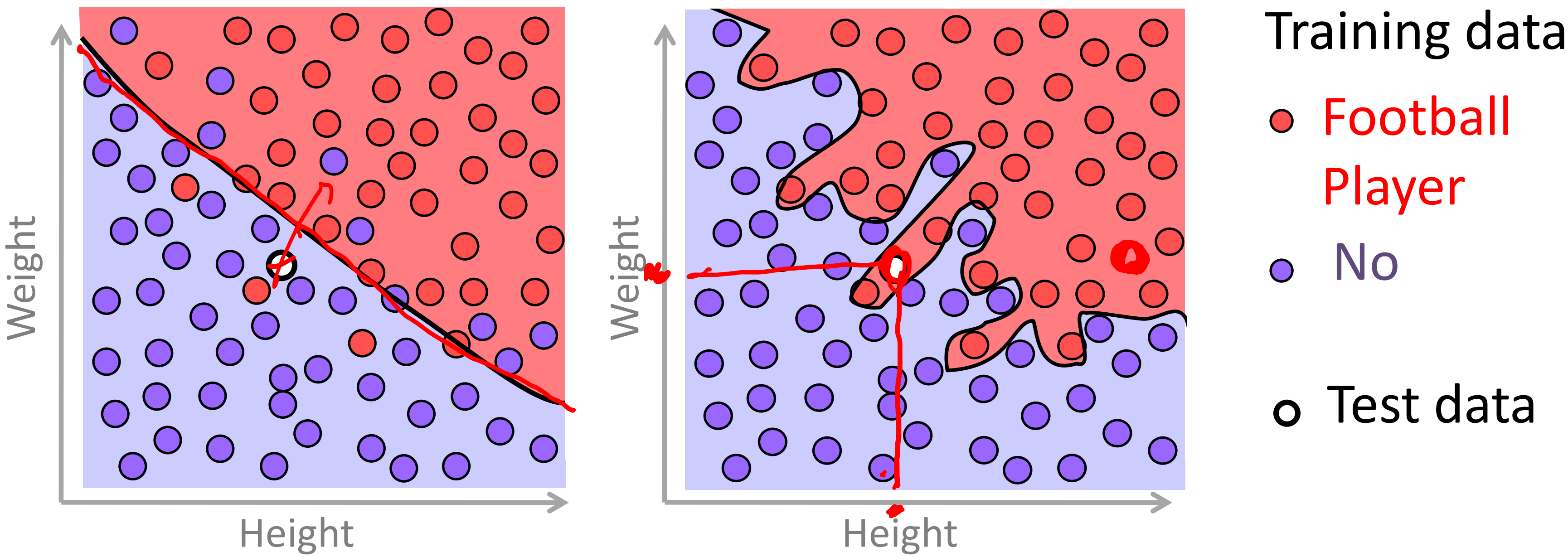
Experience



?

Performance

Training Data vs. Test Data



- A good machine learning algorithm
  - **Generalizes** aka performs well on test data
  - ~~Does not **overfit** training data~~

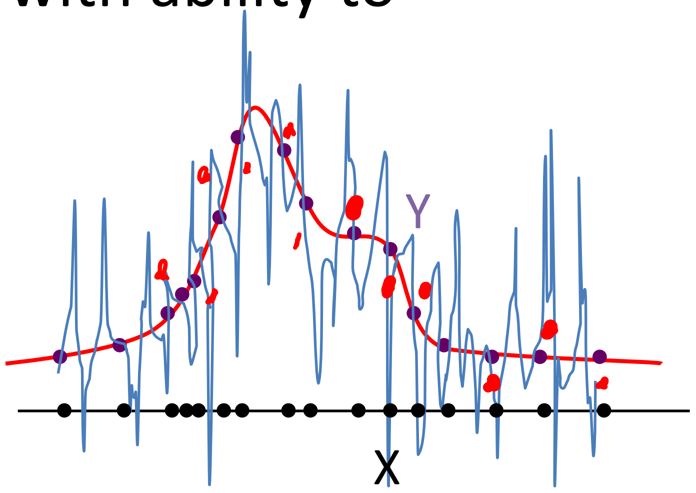
Memorizing vs. Learning

- Is it okay to **overfit** training data?
- Is it okay to **memorize** training data?

Sometimes yes (e.g. if labels are noiseless)

BUT needs to be accompanied with ability to generalize

➢ Which fit is better (Red/Blue)?



- What is learning really?

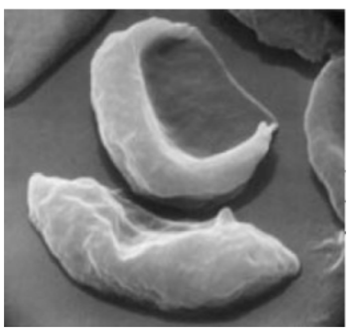
Can algorithm **generalize** aka perform well on test data

Tasks, Experience, **Performance**

Performance Measure

Performance:

$\text{loss}(Y, f(X))$ - Measure of closeness between label *Y* and prediction *f*(*X*) for test data X

X	Diagnosis, *Y*	*f*(*X*)	$\text{loss}(Y, f(X))$
	"Anemic cell"	"Anemic cell"	0
		"Healthy cell"	1

$$1_A = \begin{cases} 1 & A \\ 0 & A^c \end{cases}$$

$$\text{loss}(Y, f(X)) = 1_{\{f(X) \neq Y\}}$$ **0/1 loss**

Performance Measure

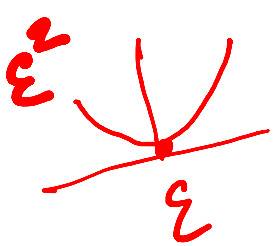
$y = mx + c$

$\frac{dy}{dx} = m$

Performance:

$\text{loss}(Y, f(X))$ - Measure of closeness between label *Y* and prediction *f*(*X*) for test data X

X	Share price, *Y*	*f*(*X*)	$\text{loss}(Y, f(X))$
Past performance, trade volume etc. as of Sept 8, 2010	"$24.50"	"$24.50"	0
		"$26.00"	1?
		"$26.10"	2?



$$\text{loss}(Y, f(X)) = (f(X) - Y)^2$$ **squared loss**

Performance Measure

For test data X, measure of closeness between label *Y* and prediction *f*(*X*)

Binary Classification $\text{loss}(Y, f(X)) = 1_{\{f(X) \neq Y\}}$ **0/1 loss**

Regression $\text{loss}(Y, f(X)) = (f(X) - Y)^2$ **squared loss**

Lets think of unsupervised tasks next.

Performance Measure

For test data X, measure how good is the learnt distribution, clustering or embedding *f*(*X*)

Learning a distribution $X \rightarrow p(X)$

Clustering $X \rightarrow C_X \in \{C_1 \cdots C_k\}$

Groups 1-10: [Jamboard_1_10]
Groups 11-20: [Jamboard_11_20]

Dimensionality reduction

$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{bmatrix} = X \rightarrow X'$, $X \in \mathbb{R}^D$, $X' \in \mathbb{R}^d$, $d < D$

➢ What performance measure would you use for each task?

Performance Measure

For test data X, measure how good is the learnt distribution, clustering or embedding *f*(*X*)

Learning a distribution $\quad x \rightarrow P(x)$

$$\sum_x \left(P(x) - P_{groundtruth}(x) \right)^2$$

(Training)
“Likelihood”

$x_1 \ldots x_n$

$$-\log \prod_{i=1}^{n} P(x_i)$$

Negative log likelihood ← loss

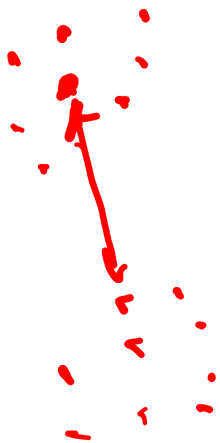
$$= \log \frac{1}{\prod_{i=1}^{n} P(x_i)}$$

Performance Measure

For test data X, measure how good is the learnt distribution, clustering or embedding *f*(*X*)

Clustering

$$\frac{\sum_{i \in C_x} dist(x, x_i)}{\sum_{j \notin C_x} dist(x, x_j)}$$



Performance Measure

For test data X, measure how good is the learnt distribution, clustering or embedding *f*(*X*)

Dimensionality reduction

$x \in \mathbb{R}^D \rightarrow x' \in \mathbb{R}^d$

~~$dist(x, x')$~~

$x' \rightarrow \tilde{x}$

$dist(x, \tilde{x})$

$\tilde{x}$ = Reconstruction of x from projection x' (discuss later how)

Glossary of Machine Learning

- Task
- Supervised learning
  - Classification
  - Regression
- Unsupervised learning
  - Learning distribution
  - Clustering
  - Dimensionality reduction/Embedding
- Input, X
- Label, Y
- Prediction, f(X)

- Experience = Training data
- Test data
- Overfitting
- Generalization
- Performance
- Likelihood
- Loss – 0/1, squared, negative log likelihood