Machine Learning

CSE 142

Xin (Eric) Wang

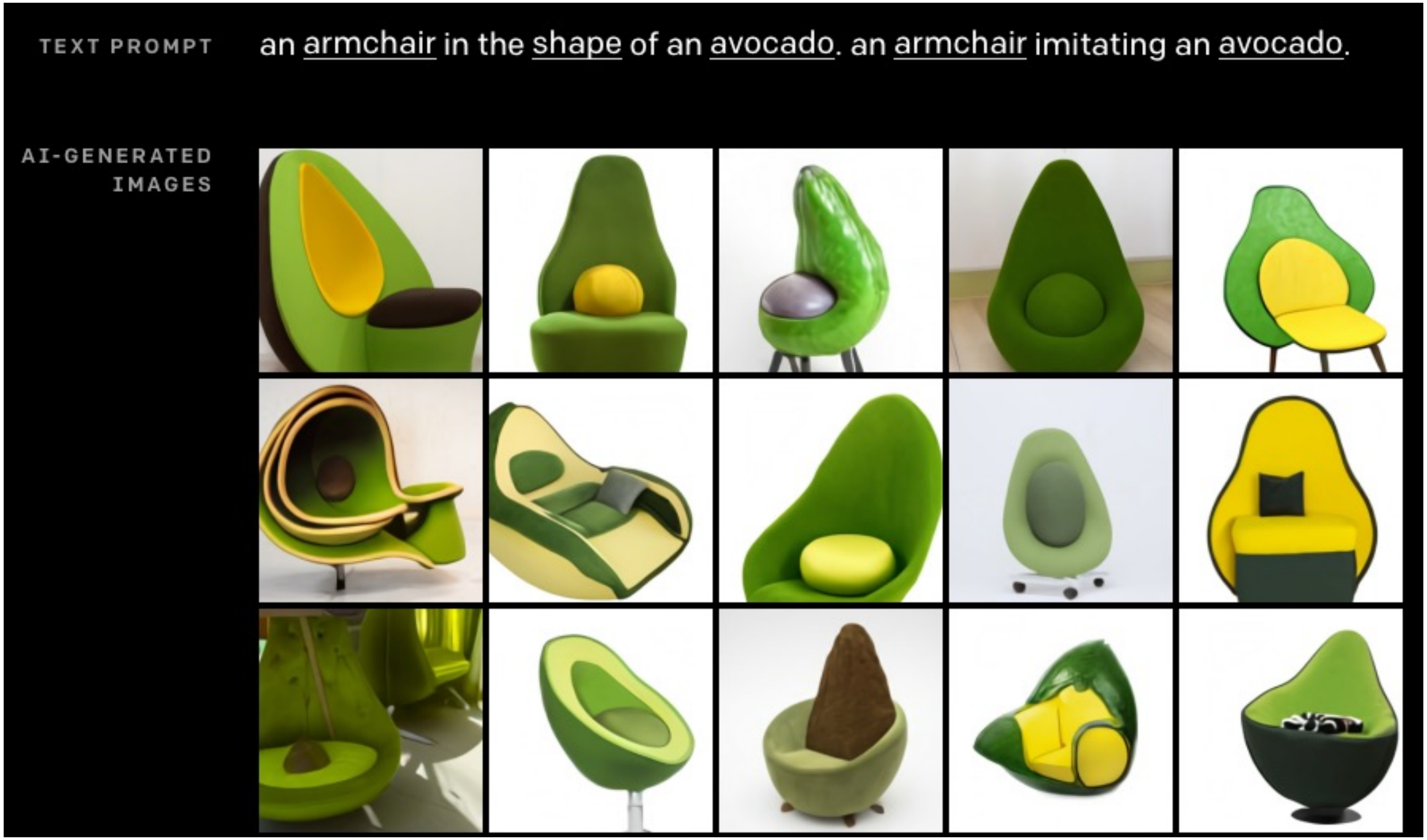
Friday, October 1, 2021

Today

- The ingredients of machine learning, Ch. 1 (cont.)
- Classification, Ch. 2

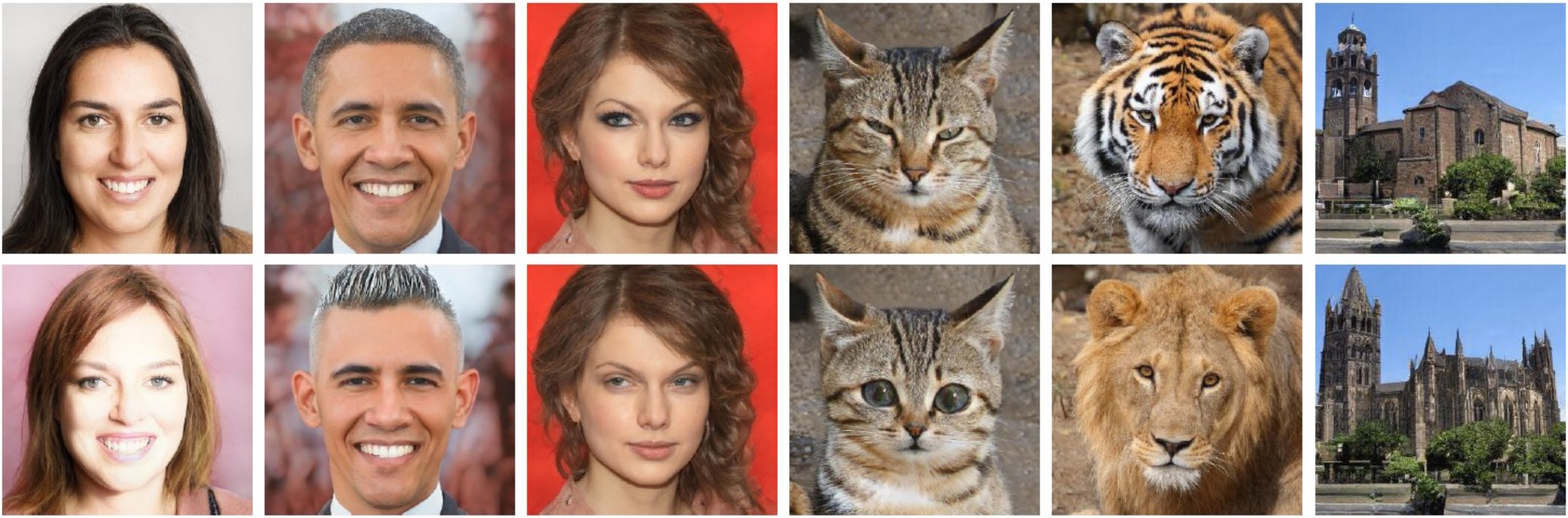
Generative Models

OpenAI's text2image generation model: https://openai.com/blog/dall-e/



Generative Models

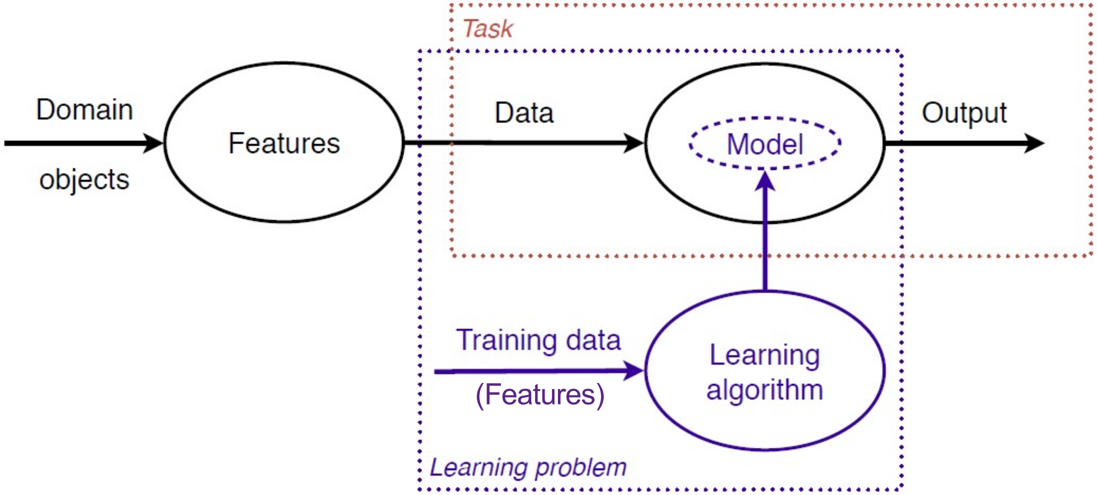
StyleCLIP: https://arxiv.org/abs/2103.17249



"Emma Stone" "Mohawk hairstyle" "Without makeup" "Cute cat" "Lion" "Gothic church"

Machine learning

Machine learning is about using the right features to build the right models that achieve the right tasks



Features – how we describe our data objects

Model – a mapping/function from data points to outputs:

$$Output = f\,(Data)$$

This is what machine learning produces!

Task – an abstract representation of the problem we want to solve

Tasks

- Classification – assign the target variable to one of N states
- Regression – assign the target variable to a real-valued (scalar or vector) function of the input
- Clustering – grouping data without prior information (unlabeled data)
- Dimensionality reduction
- Anomaly detection
- Subgroup discovery
- Association rule learning
- … …

Tasks: predictive and descriptive

- The most common ML tasks are predictive, aiming to predict/estimate a target variable from features:
  - Binary and multi-class classification: categorical target
    - Learn decision boundaries
  - Regression: numerical target
    - Learn relationship (a real-valued function) between input and output spaces

- Descriptive tasks are concerned with exploiting underlying structure in the data, finding patterns:
  - No specific problem to solve per data element
  - Goal: discover "interesting things" about the data
  - E.g., (descriptive) clustering
    - Grouping data without prior information

Quiz: Machine Learning Settings

Google Form: https://forms.gle/uKgBojuddLeX22199

Are these tasks predictive model or descriptive model? *

	Predictive model	Descriptive model
Classification	○	○
Regression	○	○
Clustering	○	○
Subgroup discovery	○	○
Association rule learning	○	○

Models

- Machine learning models can be distinguished according to their main intuition:
  - Geometric models use intuitions from geometry such as separating (hyper-)planes, linear transformations, and distance metrics
  - Probabilistic models view learning as a process of reducing uncertainty, modelled by means of probability distributions
  - Logical models are defined in terms of easily interpretable logical expressions
- Alternatively, they can be characterized by their *modus operandi*:
  - Grouping models divide the instance space (the space of possible inputs) into segments; in each segment a different (perhaps very simple) model is learned
  - Grading models learning a single, global model over the instance space

Logical models

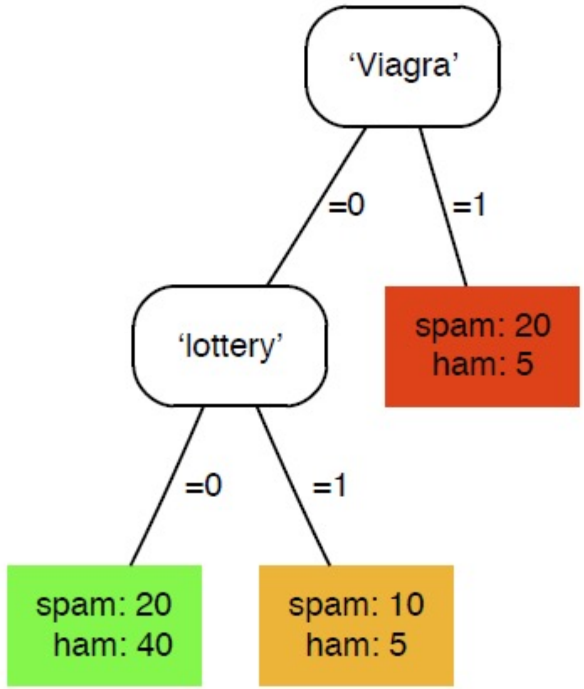
- Geometric and probabilistic models don't necessarily translate to human-understandable rules
- Logical models focus on the level of human reasoning
  - Often provide explanations for their results
- Classical AI models encapsulate logical rules and relationships for deductive reasoning

Logical models (cont.)

- Logical models in ML are often organized in tree structures – feature trees – that iteratively partition the space of all possible inputs (the *instance space*)
  - The nodes represent decisions based on feature values
  - The leaves correspond to regions of the instance space – i.e., groups of feature values
- Feature trees whose leaves are labelled with classes are called ***decision trees***

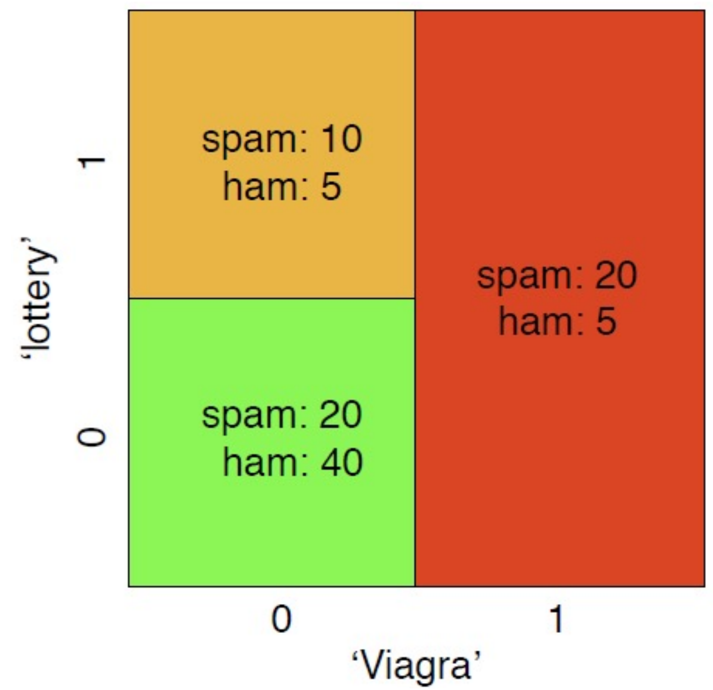
Feature trees

A feature tree for Spam vs. Ham



The partitioned instance space

Showing leaves of the feature tree

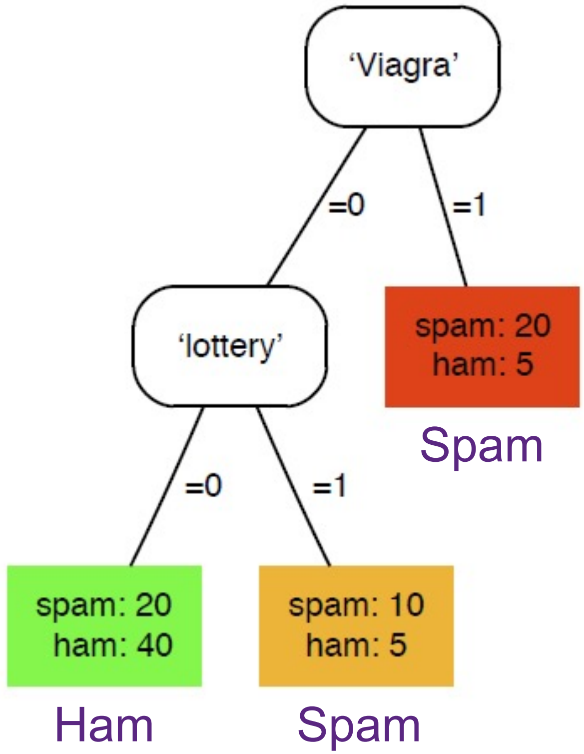


The numbers correspond to the number of emails in each bin of the training data

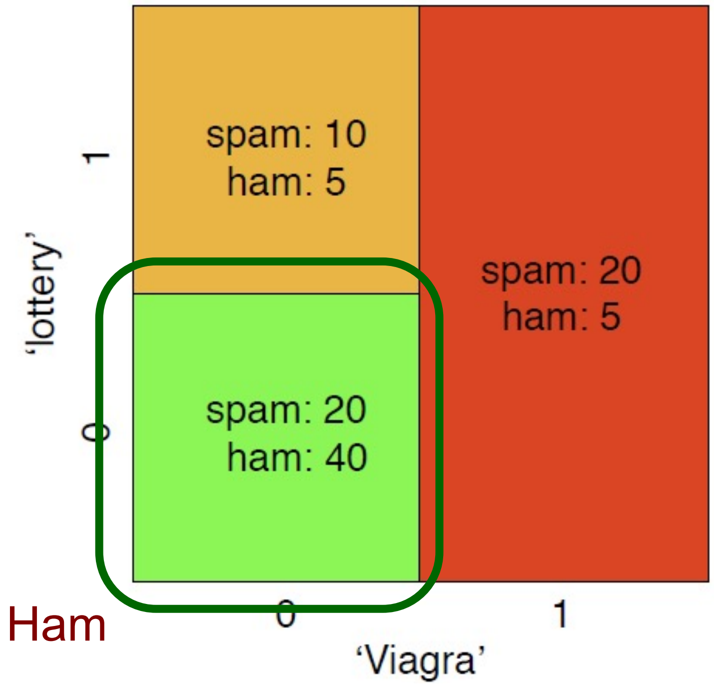
What about Viagra=1 ∧ lottery=0 ? Viagra=1 ∧ lottery=1 ?

Decision trees

A decision tree for Spam vs. Ham



The partitioned instance space



Key question for ML: How to construct a (good) decision tree from data?

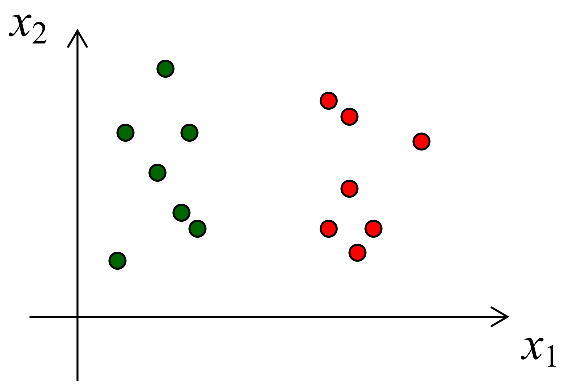
Features

- A machine learning model is only as good as its features
  - *Garbage in, garbage out!*
- Features are measurements performed on instances
  - Multiple features for an instance comprise a feature vector
  - Most often numerical, but not always
    - e.g., a feature could be "the most frequent word in the text"
- Typical feature types:
  - Boolean
  - Integers
  - Real numbers
  - Sets
- What other features might you use to classify "spam or ham"?

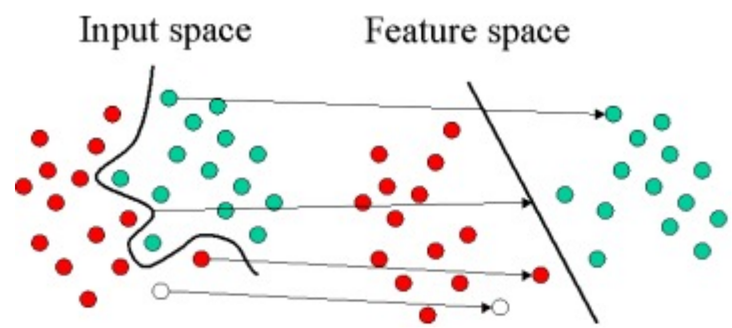
Feature construction

- In some ML problems, the features are fixed, given to you
- But in others, feature construction may be the most important part of your solution
- The "raw" features may not be optimal for the problem
  - They may have irrelevant dimensions
  - They may depend on irrelevant parameters
  - Some features may be particularly noisy (unreliable)
- We want features that:
  - Encapsulate the key similarities and differences in the data
  - Are robust to irrelevant parameters and transformations
  - Have a high signal-to-noise ratio
- Thus the first step in a ML problem is often to transform the features into a new feature space

Feature construction/transformation



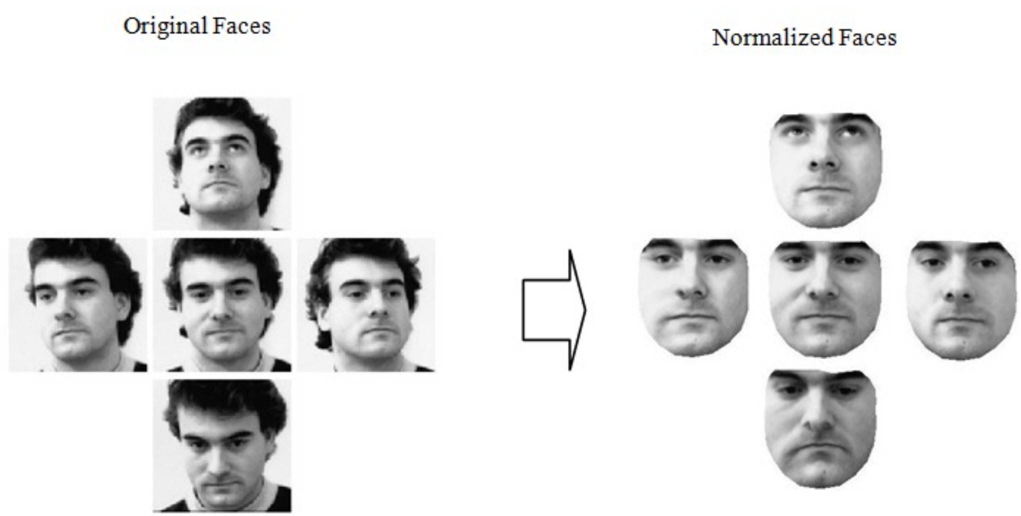
x_2 has no useful information for classifying, so transform the feature space by projecting onto the x_1 axis



Transformation (ϕ) to make the feature space linearly separable

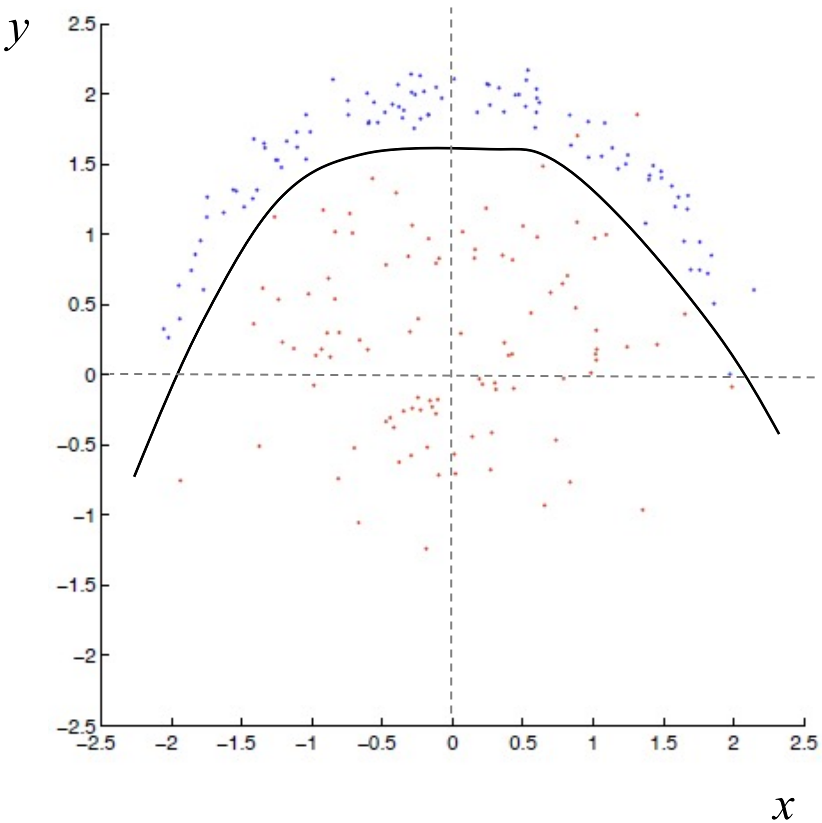
Transformation (ϕ) to make the measurements in the appropriate coordinate system

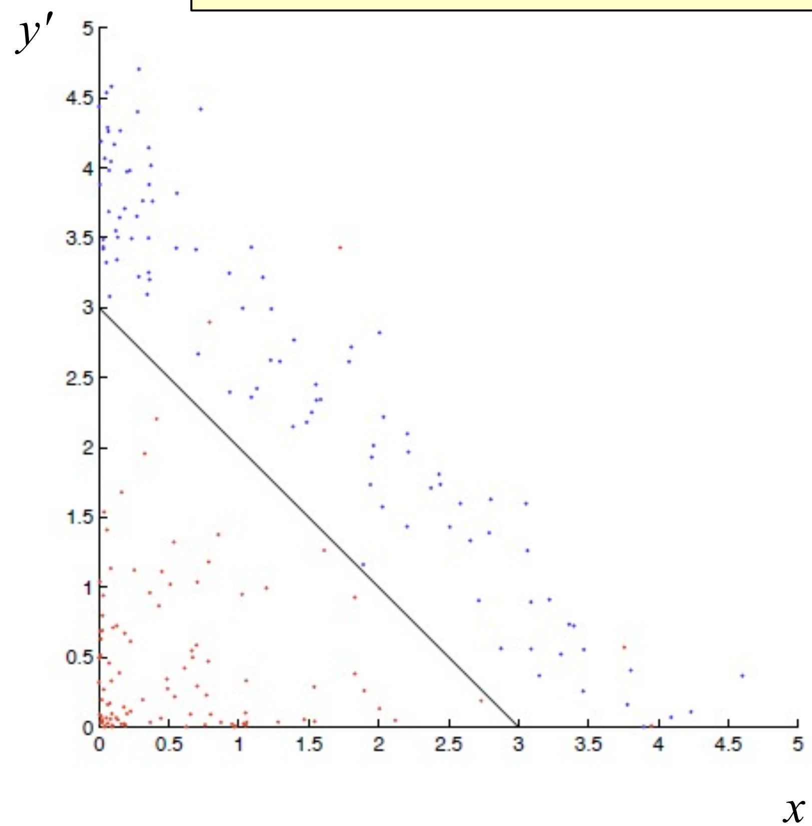
- Align the faces



Feature construction/transformation

$(x',y') = \phi(\mathrm{x}, \mathrm{y}) = (x^2, y^2)$





1. Transform features to new feature space – mapping ϕ from (x, y) to (x',y')
2. Perform linear classification

[The kernel trick – skipping for now]

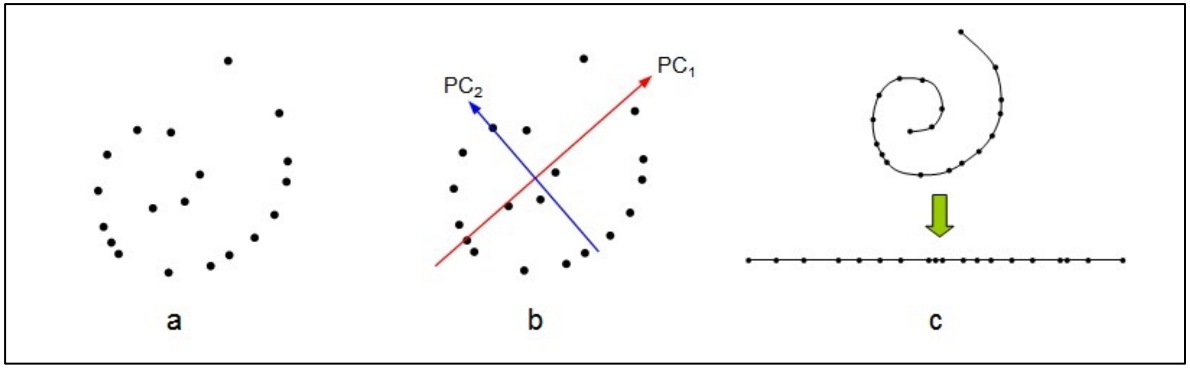
Dimensionality reduction

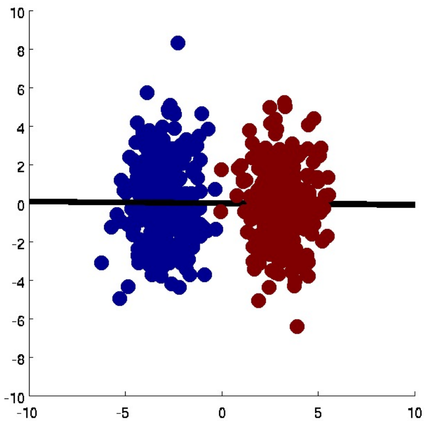
- Let's say we build a ML system to predict someone's occupation. The 12 features given to us are:
  - First name, last name, SSN, father's occupation, mother's occupation, highest educational degree, last year's salary, federal taxes paid last year, home address, model of car, miles driven last year, miles flown last year
- Some of these features may be useless, with no relevant information
- There may be redundancies – correlations among features
- Can we transform this 12-dimensional classification problem to a lower-dimensional problem?
  - Perhaps easier, computationally simpler…
- Yes, through dimensionality reduction

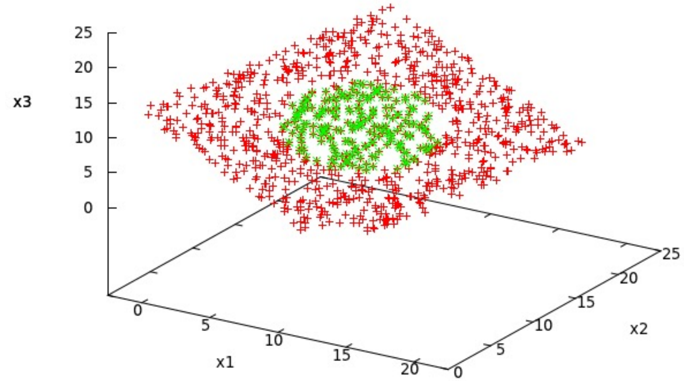
Intrinsic dimensionality

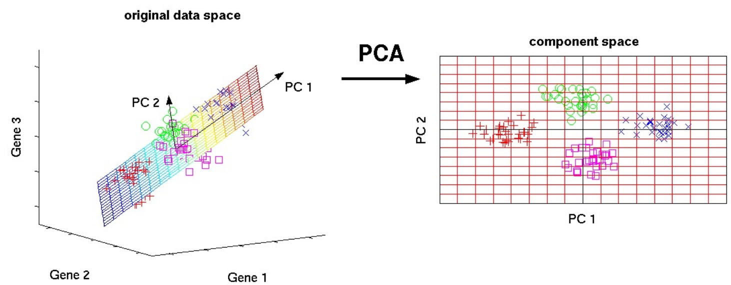
- The intrinsic dimensionality of (N-dimensional) data describes the real structure of the data, embedded in N-space
  - I.e., how many variables are needed to (minimally) represent the data?
- N-dimensional data could be intrinsically:
  - 0 dimensional – tightly clustered around a point
  - 1 dimensional – defined by a line or contour
  - 2-dimensional – defined by a plane or 2D manifold
  - M-dimensional ($M \leq N$) – defined by an M-dimensional hyperplane or M-dimensional manifold

Manifold – locally M-dimensional surface



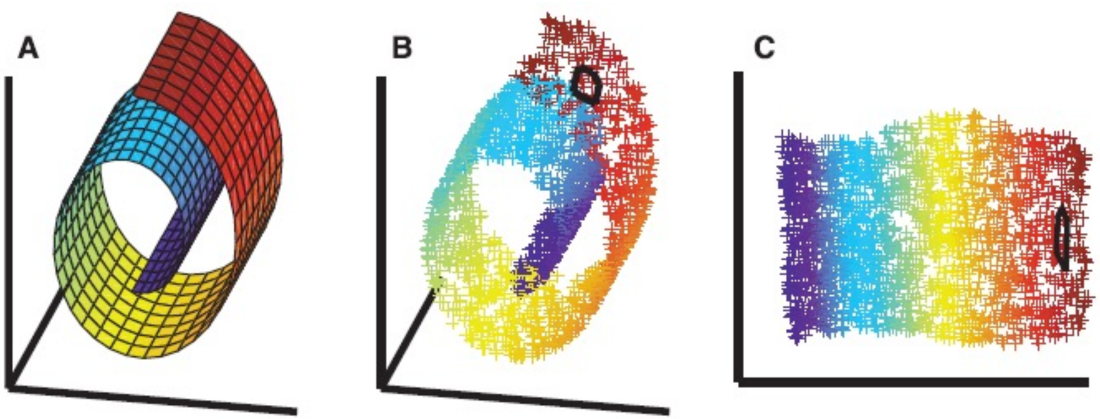


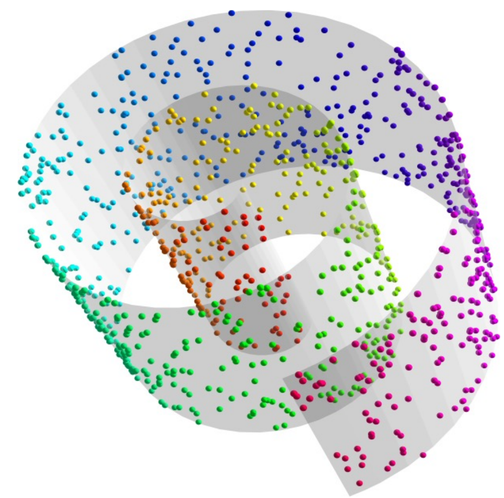




There are many dimensionality reduction methods:

- PCA
- ICA
- LDA
- LLE
- Factor analysis
- Random methods
- etc….





Summary so far…

- Key machine learning concepts
  - Core ML problem formulation
  - Important types of ML problems
  - Data sets (training, validation, test)
  - Linear classification
  - Models: generalization and overfitting
  - Models: geometric, probabilistic, logical
  - Distance measures
  - Tasks: predictive and descriptive
  - Features
  - The curse of dimensionality; intrinsic dimensionality
- Next:
  - Classification
    - Formulation, assessment, methods

Classification

Chapter 2 in the textbook

You be the classifier – learn the concept

Training data:

Engaged	Yesterday
Dedicated	Running
Devotion	Play
Work	Giraffe
Ground	Coupon
Live	Russia
Fathers	Coffee
Advanced	Ceramic
Class 1	**Class -1**

Verification data:

Party	Leisure
Restaurant	Power
Equal	Resting
Kitten	Eating
Proposition	Nation
Great	Minus
Computer	Kiss
Court	Field
Honored	Slug

- Form a hypothesis (model, function)…
- Let's test it on a verification data set:
  - https://forms.gle/6i5XTeiyjYrwitRq8

You be the classifier – learn the concept

Training data:

Engaged	Yesterday
Dedicated	Running
Devotion	Play
Work	Giraffe
Ground	Coupon
Live	Russia
Fathers	Coffee
Advanced	Ceramic
Score	Wedding
Four	Job
Seven	Sky
Continent	Phoenix
Nation	City
War	Peace
Consecrate	Marrow

Verification data:

Party	Leisure
Restaurant	Power
Equal	Resting
Kitten	Eating
Proposition	Nation
Great	Minus
Computer	Kiss
Court	Field
Honored	Slug

Try again with more training data…

Now we're finished training…!

You be the classifier – learn the concept

Training data:

Engaged	Yesterday
Dedicated	Running
Devotion	Play
Work	Giraffe
Ground	Coupon
Live	Russia
Fathers	Coffee
Advanced	Ceramic
Score	Wedding
Four	Job
Seven	Sky
Continent	Phoenix
Nation	City
War	Peace
Consecrate	Marrow

Test data:

Grief	Cause
Years	Camera
Forth	Pool
Endure	Freedom
California	Flashlight
Random	History
Battle	Nobly
Esteemed	Birth
Strike	People
Union	Newspaper
Resting	Perish
Education	Saga
Dead	Money
Earth	Moon

You be the classifier – learn the concept

Training data:

Engaged	Yesterday
Dedicated	Running
Devotion	Play
Work	Giraffe
Ground	Coupon
Live	Russia
Fathers	Coffee
Advanced	Ceramic
Score	Wedding
Four	Job
Seven	Sky
Continent	Phoenix
Nation	City
War	Peace
Consecrate	Marrow

Test data:

Grief	Cause
Years	Camera
Forth	Pool
Endure	Freedom
California	Flashlight
Random	History
Battle	Nobly
Esteemed	Birth
Strike	People
Union	Newspaper
Resting	Perish
Education	Saga
Dead	Money
Earth	Moon

The Gettysburg Address

Four score and seven years ago our fathers brought forth on this continent, a new nation, conceived in Liberty, and dedicated to the proposition that all ~~men~~(humans) are created equal.

Now we are engaged in a great civil war, testing whether that nation, or any nation so conceived and so dedicated, can long endure. We are met on a great battle-field of that war. We have come to dedicate a portion of that field, as a final resting place for those who here gave their lives that that nation might live. It is altogether fitting and proper that we should do this.

But, in a larger sense, we can not dedicate -- we can not consecrate -- we can not hallow -- this ground. The brave men, living and dead, who struggled here, have consecrated it, far above our poor power to add or detract. The world will little note, nor long remember what we say here, but it can never forget what they did here. It is for us the living, rather, to be dedicated here to the unfinished work which they who fought here have thus far so nobly advanced. It is rather for us to be here dedicated to the great task remaining before us -- that from these honored dead we take increased devotion to that cause for which they gave the last full measure of devotion -- that we here highly resolve that these dead shall not have died in vain -- that this nation, under God, shall have a new birth of freedom -- and that government of the people, by the people, for the people, shall not perish from the earth.

Abraham Lincoln
November 19, 1863

You be the classifier – learn the concept

A few observations/questions:

- What if the word "iPad" were accidentally included in the training positives? If the word "ago" were accidentally included in the training negatives?
  - Machine learning has to allow for the possibility of noisy training data
    - E.g., *uncertainty* of training labels
- What if you had never heard of the Gettysburg Address?
  - This task assumed that the learner (you) had certain semantic knowledge at your disposal – this knowledge allowed you to extract meaningful features from the data (the word instances)
  - Acquiring such semantic knowledge is a very hard problem!
  - Without it, no classifier could do well on this task
- What is the *chance* (random) *performance* in this task?

NEXT

Task	*Label space*	*Output space*	*Learning problem*
Classification	$\mathscr{L} = \mathscr{C}$	$\mathscr{Y} = \mathscr{C}$	learn an approximation $\hat{c} : \mathscr{X} \to \mathscr{C}$ to the true labelling function c
Scoring and ranking	$\mathscr{L} = \mathscr{C}$	$\mathscr{Y} = \mathbb{R}^{\lvert\mathscr{C}\rvert}$	learn a model that outputs a score vector over classes
Probability estimation	$\mathscr{L} = \mathscr{C}$	$\mathscr{Y} = [0,1]^{\lvert\mathscr{C}\rvert}$	learn a model that outputs a probability vector over classes
Regression	$\mathscr{L} = \mathbb{R}$	$\mathscr{Y} = \mathbb{R}$	learn an approximation $\hat{f} : \mathscr{X} \to \mathbb{R}$ to the true labelling function f

Don’t forget the assigned reading (at least Ch. 2)!

Some key terms in classification

- Task, model, features, instances
- Feature space, instance space, label space, output space
- Training set of labeled instances
- Instance noise, label noise
- Labeling function

- Set theory (discrete math) terms
  - Set, null set, power set
  - Intersection, union, difference, complement, cardinality
  - Cartesian product
  - Set relations
  - Properties of relations (reflexive, symmetric, antisymmetric, transitive, total
  - Equivalence relation, equivalence class, partition