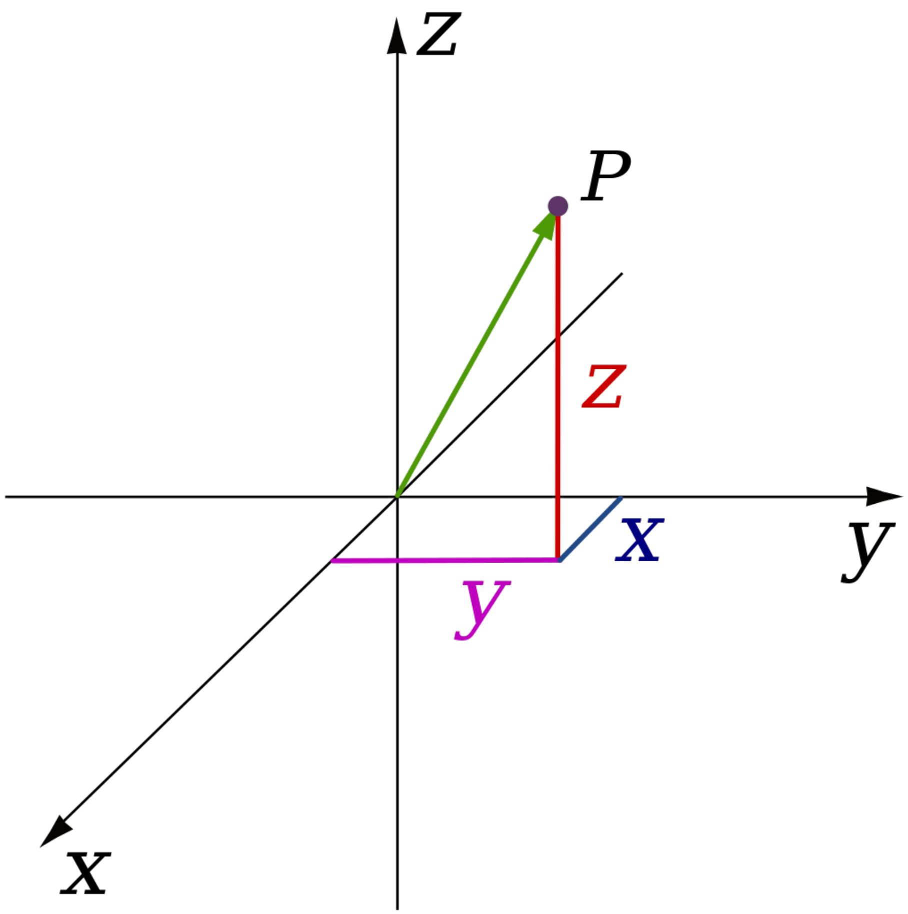
Spaces and Points

Lecture 3a

CS3400 Machine Learning

Space



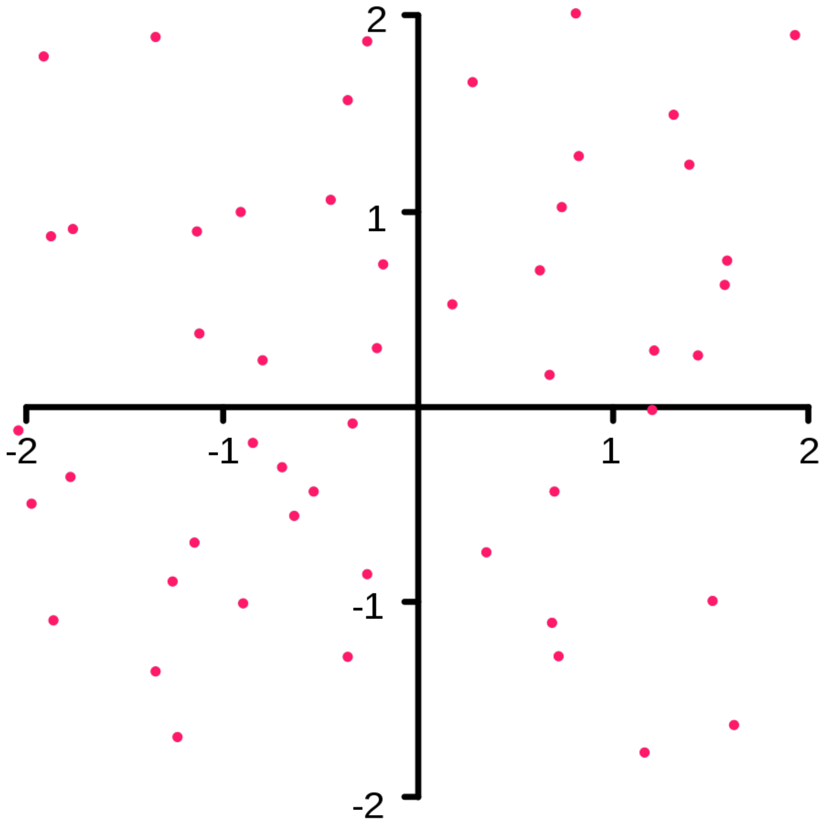
- Coordinate system
- Has a specified number of dimensions
- Examples:
  - 2D – x, y
  - 3D – x, y, z
- Not restricted to 2 or 3 dimensions – in this course we'll see examples of high-dimensional spaces!

Geographic Coordinate System

- Latitude and longitude form a 2D coordinate system or space
- Used to describe locations across the globe
- MSOE: 43.041070, -87.909420
- Miller Park: 43.011790, -87.967780
- Margate, FL: 26.242530, -80.204920

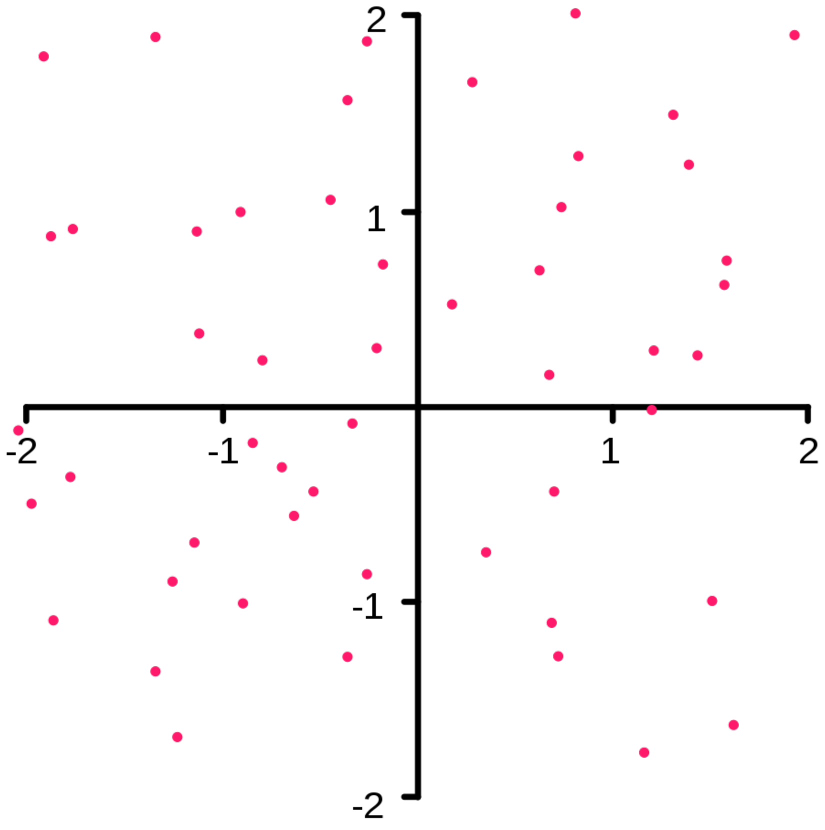
https://www.ergosum.co/cooking-across-latitudes/

Points



- A point is a 0-dimensional geometric object (has no volume)
- Represented as a tuple of N numbers (one for each dimension in the space)
  - p = (x, y)
  - p = (x, y, z)

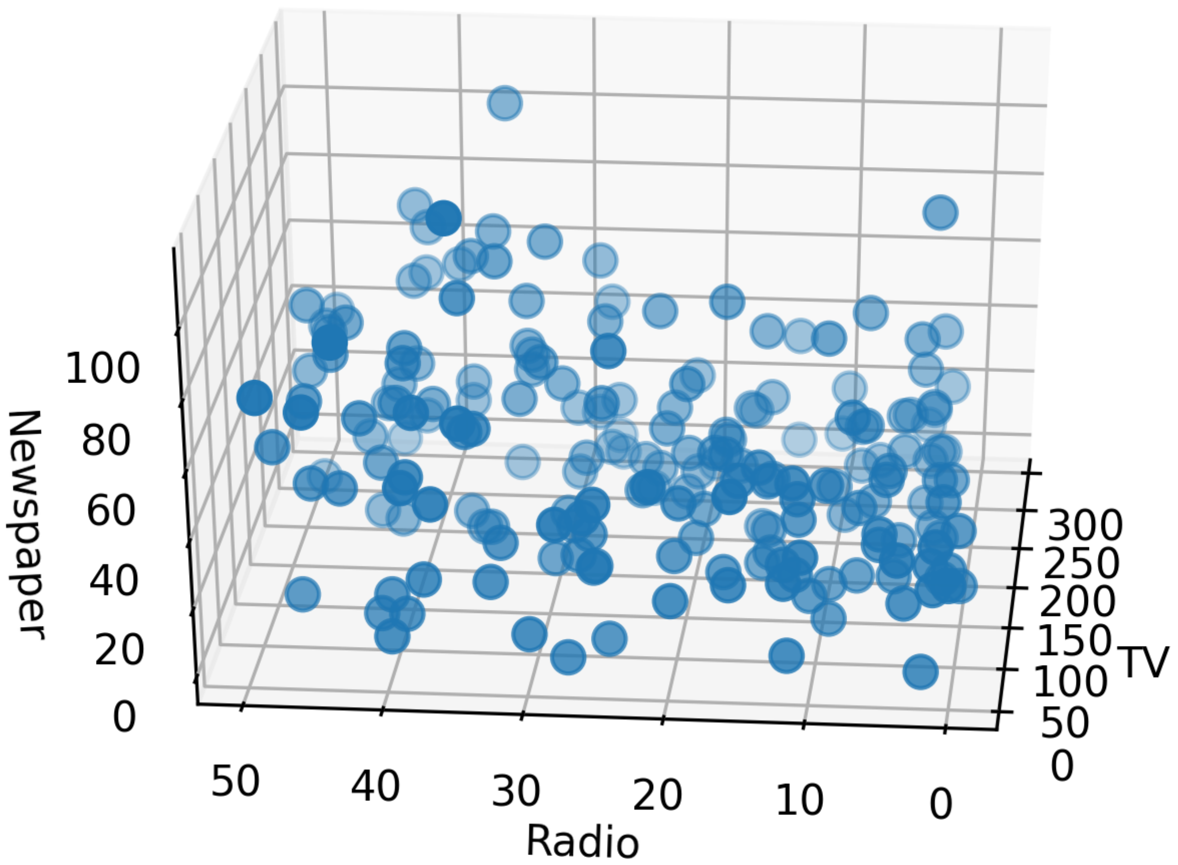
Points



Examples of 2D points:

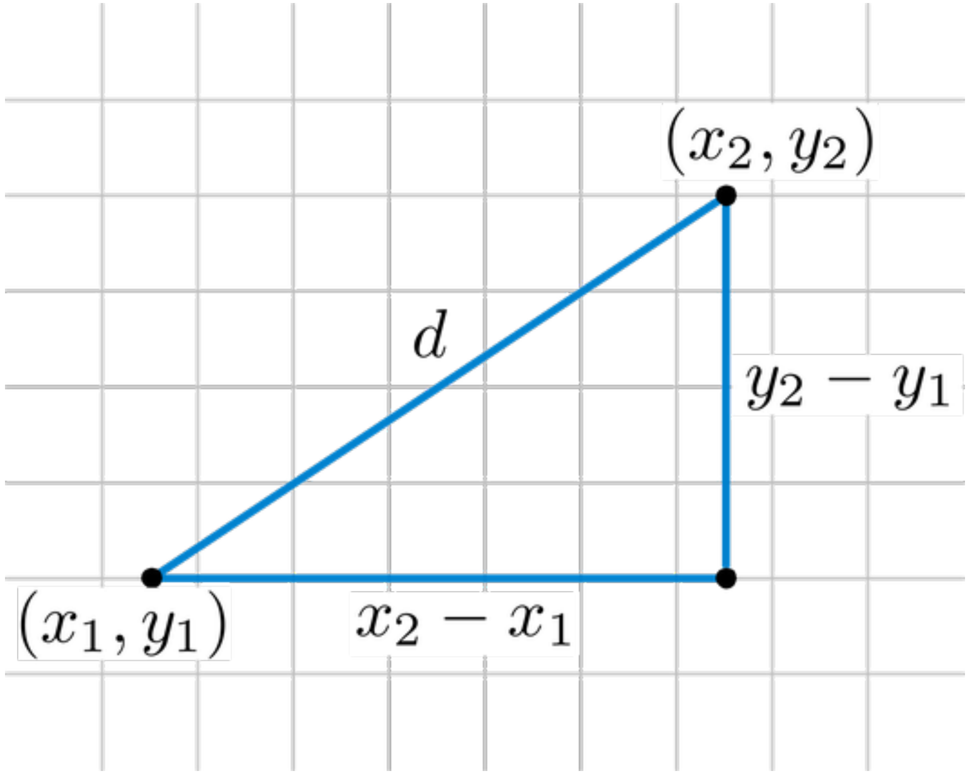
- (0.75, 0.2)
- (1.25, 0)
- (1.25, -1.75)

Examples of 3D Points



- Each point is the client of an advertising company
- The coordinates refer to the number advertisements they bought
  - TV advertisements
  - radio advertisements
  - newspaper advertisements

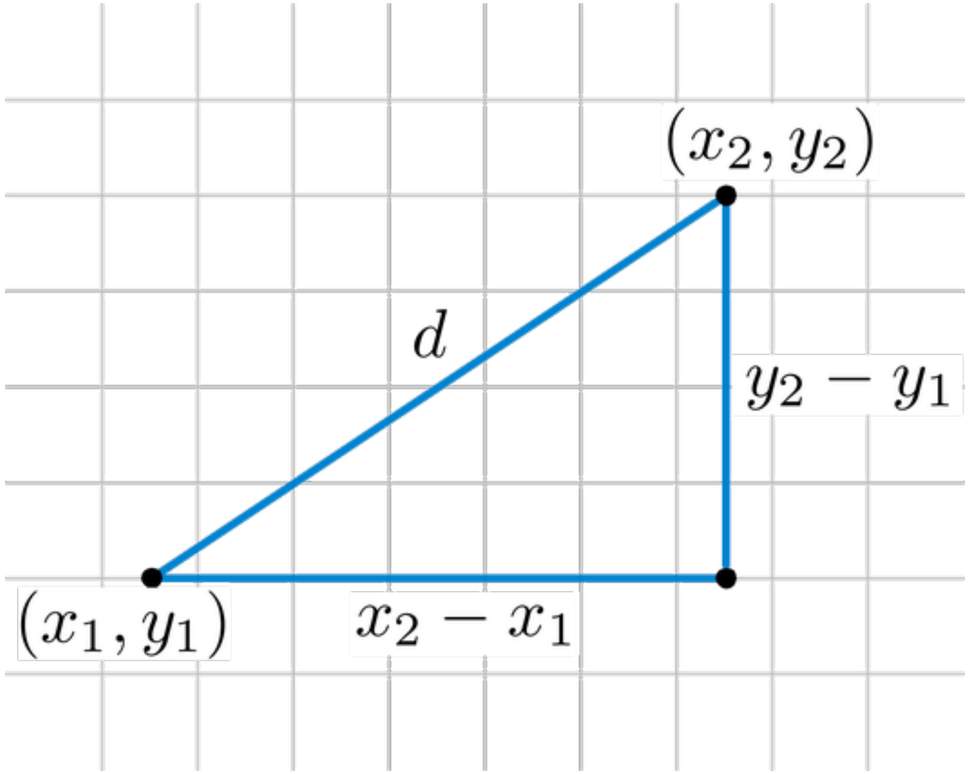
Calculating Distance



http://rosalind.info/glossary/euclidean-distance/

- We can calculate the distance between two points
- A distance function is any function d(p0, p1) that satisfies three properties:
    - d(p0, p0) = 0 – the distance of any point and itself is 0
    - d(p0, p1) >= 0 – distances cannot be negative
    - d(p0, p1) <= d(p0, p2) + d(p2, p1) –triangle inequality

Euclidean Distance



You are probably familiar with Euclidean distance:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

for the 2D points (x1, y1) and (x2, y2) and

$$\mathrm{d} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

for the 3D points (x1, y1, z1) and (x2, y2, z2)

Euclidean distance can be extended to any number of dimensions

http://rosalind.info/glossary/euclidean-distance/

Spaces in This Course

In this course, we're going to talk about two main types of spaces:

- Feature spaces
  - Features are variables containing measurements of objects (e.g., length)
  - A feature space is the collection of all features describing objects
  - Each point in our data set is a point in the feature space
- Parameter spaces
  - Models have variables called parameters
  - The parameters form a space
  - Each fitted model is a unique combination of the parameter values that is a distinct point in the parameter space
  - The parameter space represents all possible models

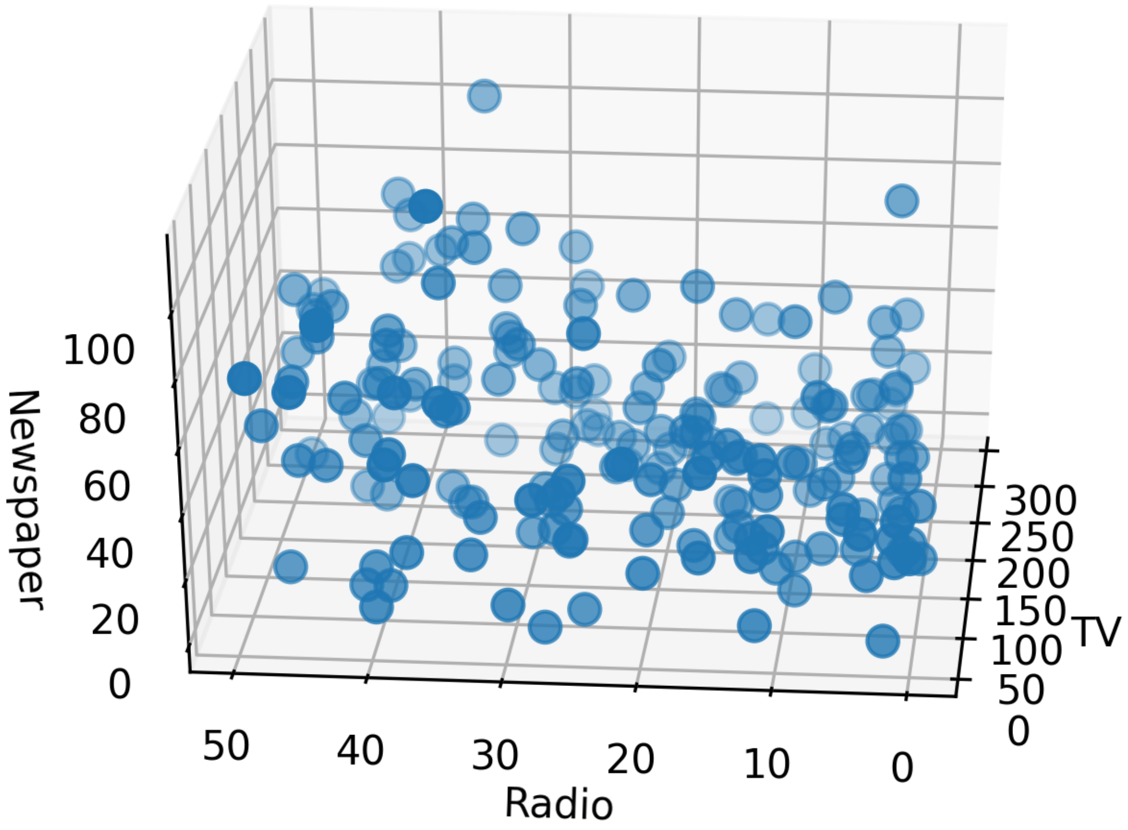
Advertising Data Set

- Each record corresponds to a single company that is a client of an advertising agency
- Response: Amount of sales
- Predictors (features):
  - Amount of TV advertisements
  - Amount of radio advertisements
  - Amount of newspaper advertisements

	TV	radio	newspaper	sales
0	230.1	37.8	69.2	22.1
1	44.5	39.3	45.1	10.4
2	17.2	45.9	69.3	9.3
3	151.5	41.3	58.5	18.5
4	180.8	10.8	58.4	12.9

Feature Space

- The data set has 3 features
  - TV advertisements
  - radio advertisements
  - newspaper advertisements
- The feature space therefore has 3 dimensions
- Note that the feature space does not include the output variable!
- These features describe each object as a point in the feature space



Linear Regression

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

where $\hat{y}$ is the predicted value, *p* is the number of features, x_i are the features, and β_i are the feature weights.

Advertising Data

$$\widehat{sales} = \beta_0 + \beta_1 TV + \beta_2 radio + \beta_3 newspaper$$

where $\widehat{sales}$ is the predicted value, TV, radio, and newspaper are the features, and β_i are the feature weights.

Fitted Model

$$\widehat{sales} = 0.0874 + 0.0530 \cdot TV + 0.2215 \cdot radio + 0.0162 \cdot newspaper$$

where $\widehat{sales}$ is the predicted value, TV, radio, and newspaper are the features, and β_i are the feature weights.

Parameter Space

- The model has four 4 parameters (β_0 - β_3)
- The parameter space has 4 dimensions
- The parameter values describe each model as a point in the parameter space

Parameter Space

- We'll use the parameter space to search for model parameters that optimize a model for a given data set
- We'll use a "cost" function to measure the fit of a set of parameters
- The parameters that give the lowest cost will be give us the best model