Tuesday · September 7, 2021

Semantics: What Is Meaning?

Yale

LING 380/780

Neural Network Models of Linguistic Structure

Is a burrito a sandwich?

Peanut Butter and Jelly Sandwich


Grinder Sandwich

Gyro Sandwich


Smørrebrod
Scandinavian Open-Faced Sandwich

White City Shopping Center v. Panera Bread

- The White City Shopping Center in Shrewsbury, MA, gave Panera Bread the exclusive right to sell sandwiches.
- Panera sued White City when it started negotiating a contract with Qdoba Mexican Eats.
- They argued that burritos were sandwiches.

Is White City Shopping Center guilty?

Course Logistics

- Official textbooks for this course:
  - Course Notes by Sophie Hao
  - *Dive into Deep Learning* (D2L) by Aston Zhang, Zack C. Lipton, Mu Li, and Alex J. Smola
- Readings posted in “Modules” tab of Canvas
- Office Hours:
  - Bob: Mondays from 9:30 AM to 11:30 AM at Dow Hall 200
  - Sophie: Wednesdays from 9:30 AM to 11:30 AM on Zoom
- First assignment released this week

What is meaning?

What Is Meaning?

Leonard Bloomfield

- Sterling Professor of Linguistics at Yale, 1940–1949
- Major figure in American Structuralism
- Studied Algonquian, Austronesian, and Indo-European languages

What Is Meaning?

“We can define the meaning of a speech-form accurately when this meaning has to do **with some matter of which we possess scientific knowledge** … but we have no precise way of defining words like *love* or *hate*, which concern situations that **have not been accurately classified** – and these latter are in the great majority.

What Is Meaning?

“Moreover, even where we have some scientific (that is, **universally recognized and accurate**) classification, we often find that the meanings of a language do not agree with this classification. The whale in German is called a ‘fish’: *Walfisch* … and the bat a ‘mouse’: *Fledermaus* ….

What Is Meaning?

"Physicists view the color-spectrum as a continuous scale of light-waves of different lengths, ranging from 40 to 72 hundred-thousandths of a millimeter, but languages mark off different parts of this scale **quite arbitrarily and without precise limits**, in the meanings of such color-names as *violet*, *blue*, *green*, *yellow*, *orange*, *red*, and the color-names of different languages do not embrace the same gradations.

What Is Meaning?

“The kinship of persons seems a simple matter, but terminologies of kinship that are used in various languages are extremely hard to analyze.

“The statement of meanings is therefore the weak point in language-study, and will remain so until human knowledge advances very far beyond its present state.”

Leonard Bloomfield
Language (1933), Chapter 9

Structuralist Semantics

According to the structuralist worldview, words have a **precise** and **technical** meaning that can only be uncovered by **experts** through careful study.

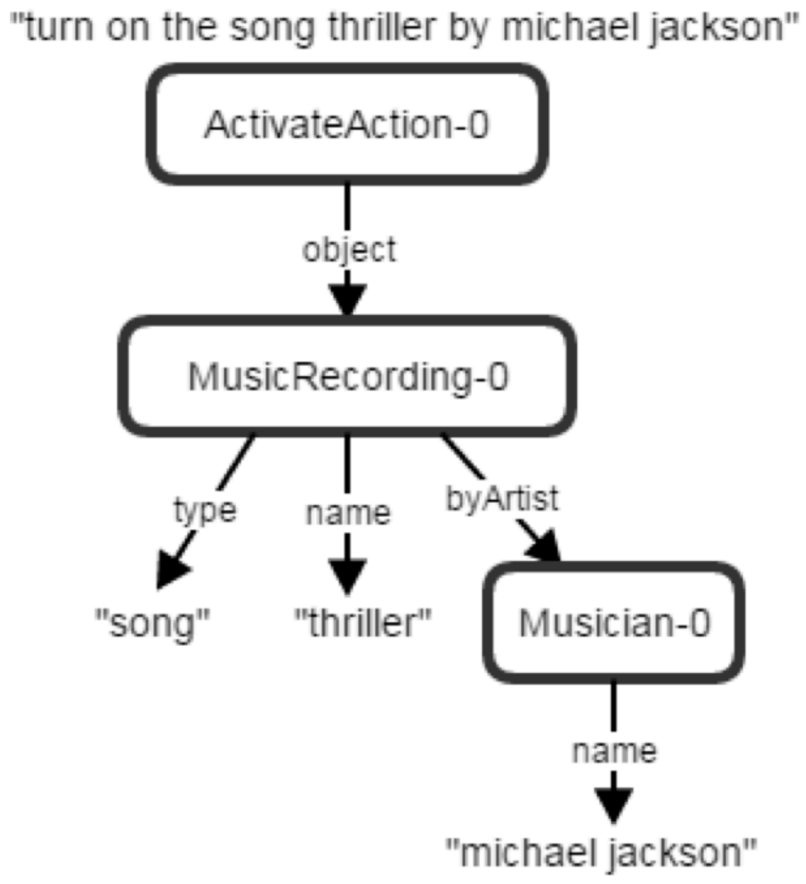
What is a sandwich, according to Bloomfield?

THE SANDWICH ALIGNMENT CHART

	INGREDIENT PURIST (Must have classic sandwich toppings: meat, cheese, lettuce, condiments, etc.)	INGREDIENT NEUTRAL (Can contain a broader scope of savoury ingredients)	INGREDIENT REBEL (Can contain literally any food products sandwiched together)
STRUCTURE PURIST (A sandwich must have a classic sandwich shape: two pieces of bread/baked product, with toppings in between)	**HARDLINE TRADITIONALISTS**  "A BLT is a sandwich."	**STRUCTURAL PURIST, INGREDIENT NEUTRAL**  "A chip butty is a sandwich."	**STRUCTURAL PURIST, INGREDIENT REBEL**  "Ice cream between waffles is a sandwich."
STRUCTURE NEUTRAL (The container must be on either side of the toppings, but not necessarily two separate pieces)	**STRUCTURAL NEUTRAL, INGREDIENT PURIST**  "A sub is a sandwich."	**TRUE NEUTRAL**  "A hot dog is a sandwich."	**STRUCTURAL NEUTRAL, INGREDIENT REBEL**  "An ice cream taco is a sandwich."
STRUCTURE REBEL (Can contain any food enveloped in any way by a containing food)	**STRUCTURAL REBEL, INGREDIENT PURIST**  "A chicken wrap is a sandwich."	**STRUCTURAL REBEL, INGREDIENT NEUTRAL**  "A burrito is a sandwich."	**RADICAL SANDWICH ANARCHY**  "A Pop-Tart is a sandwich."

Ontologies

- A giant database of world knowledge.
- WordNet Ontology
- Amazon Alexa Ontology



The Verdict

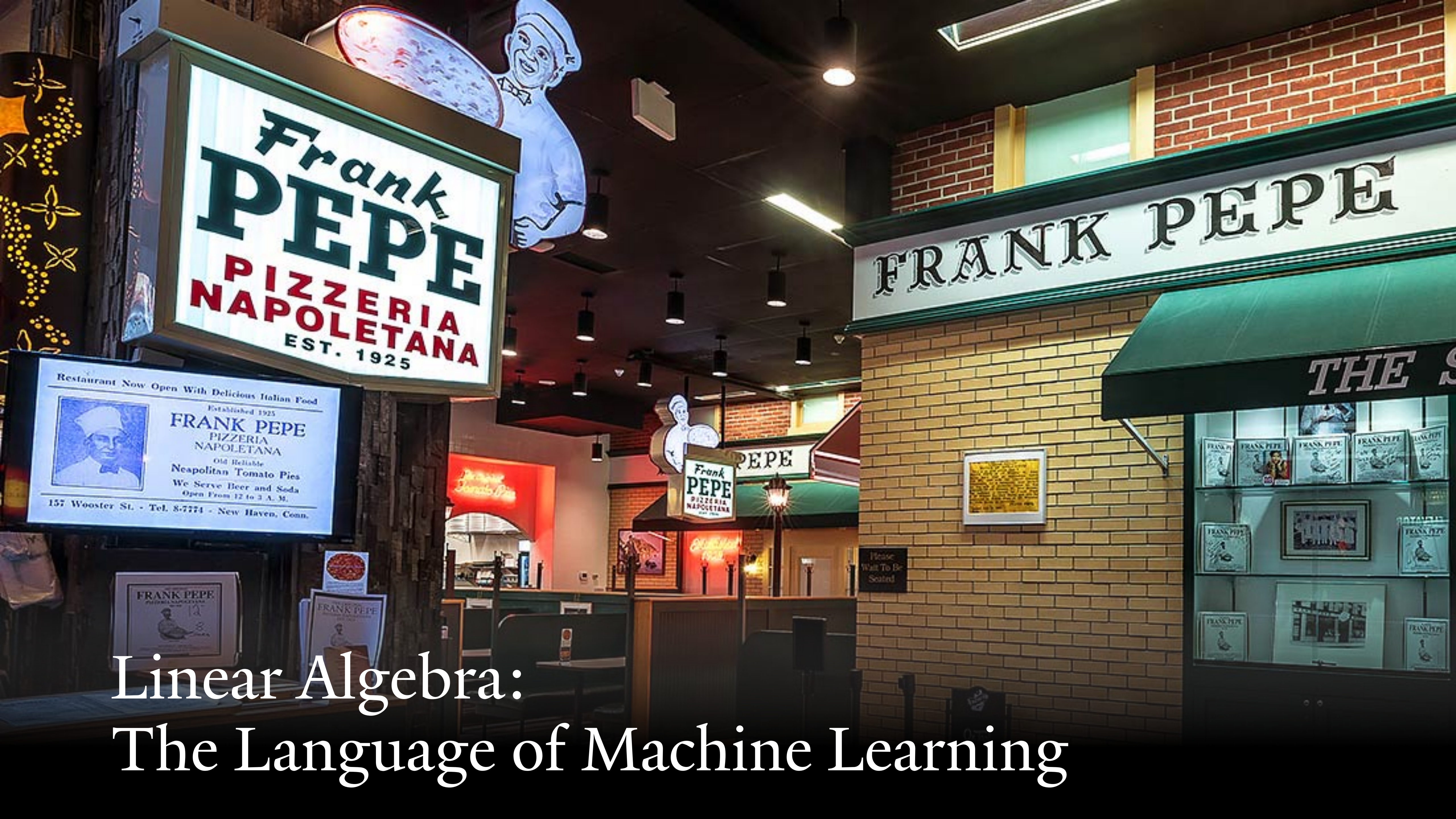
"Given that the term 'sandwiches' is not ambiguous and [Panera's] Lease does not provide a definition of it, this court applies the **ordinary meaning** of the word. The **New Webster Third International Dictionary** describes a 'sandwich' as 'two thin pieces of bread, usually buttered, with a thin layer (as of meat, cheese, or savory mixture) spread between them.' …

The Verdict

"Under this definition and **as dictated by common sense**, this court finds that the term 'sandwich' **is not commonly understood to include** burritos, tacos, and quesadillas, which are typically made with a single tortilla and stuffed with a choice filling of meat, rice, and beans. As such, there is no viable legal basis for barring White City from leasing to [Qdoba]."

Judge Jeffrey A. J. Locke

White City Shopping Center v. Panera Bread (2006), Opinion



Linear Algebra:
The Language of Machine Learning

Basic Linear Algebra

Matrices

$$\boldsymbol{A} = \begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,1} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m,1} & A_{m,2} & \cdots & A_{m,n} \end{bmatrix} \in \mathbb{R}^{m\times n}$$

Basic Linear Algebra

Row Indexing

$$\boldsymbol{A}_{i,:} = [A_{i,1} \quad A_{i,2} \quad \cdots \quad A_{i,n}]$$

Column Indexing

$$\boldsymbol{A}_{:,j} = \begin{bmatrix} A_{1,j} \\ A_{2,j} \\ \vdots \\ A_{m,j} \end{bmatrix}$$

Basic Linear Algebra

Matrix Slicing

- $\boldsymbol{A}_{i:j,k:l}$ = rows i through j, columns k through l
- $\boldsymbol{A}_{:j,k:l}$ = rows 1 through j, columns k through l
- $\boldsymbol{A}_{:j,k:}$ = rows 1 through j, columns l through n
- $\boldsymbol{A}_{i,k:l}$ = row i, columns k through l

Basic Linear Algebra

Column Vectors

$$\boldsymbol{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{R}^n$$

Basic Linear Algebra

Row Vectors

$$\boldsymbol{v}^\top = [v_1 \quad v_2 \quad \cdots \quad v_m] \in \mathbb{R}^{1\times m}$$

Basic Linear Algebra

Matrix Addition

$$\boldsymbol{A} + \boldsymbol{B} = \begin{bmatrix} A_{1,1} + B_{1,1} & A_{1,2} + B_{1,2} & \cdots & A_{1,n} + B_{1,n} \\ A_{2,1} + B_{2,1} & A_{2,2} + B_{2,2} & \cdots & A_{2,n} + B_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m,1} + B_{m,1} & A_{m,2} + B_{m,2} & \cdots & A_{m,n} + B_{m,n} \end{bmatrix}$$

Basic Linear Algebra

Matrix Subtraction

$$\boldsymbol{A} - \boldsymbol{B} = \begin{bmatrix} A_{1,1} - B_{1,1} & A_{1,2} - B_{1,2} & \cdots & A_{1,n} - B_{1,n} \\ A_{2,1} - B_{2,1} & A_{2,2} - B_{2,2} & \cdots & A_{2,n} - B_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m,1} - B_{m,1} & A_{m,2} - B_{m,2} & \cdots & A_{m,n} - B_{m,n} \end{bmatrix}$$

Basic Linear Algebra

Hadamard Product

$$\boldsymbol{A} \odot \boldsymbol{B} = \begin{bmatrix} A_{1,1}B_{1,1} & A_{1,2}B_{1,2} & \cdots & A_{1,n}B_{1,n} \\ A_{2,1}B_{2,1} & A_{2,2}B_{2,2} & \cdots & A_{2,n}B_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m,1}B_{m,1} & A_{m,2}B_{m,2} & \cdots & A_{m,n}B_{m,n} \end{bmatrix}$$

Basic Linear Algebra

Hadamard Quotient

$$\frac{\boldsymbol{A}}{\boldsymbol{B}} = \begin{bmatrix} A_{1,1}/B_{1,1} & A_{1,2}/B_{1,2} & \cdots & A_{1,n}/B_{1,n} \\ A_{2,1}/B_{2,1} & A_{2,2}/B_{2,2} & \cdots & A_{2,n}/B_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m,1}/B_{m,1} & A_{m,2}/B_{m,2} & \cdots & A_{m,n}/B_{m,n} \end{bmatrix}$$

Basic Linear Algebra

Scalar Multiplication/Division

$$c\boldsymbol{A} = \begin{bmatrix} cA_{1,1} & cA_{1,2} & \cdots & cA_{1,n} \\ cA_{2,1} & cA_{2,2} & \cdots & cA_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ cA_{m,1} & cA_{1,1} & \cdots & cA_{m,n} \end{bmatrix}$$

$$\frac{\boldsymbol{A}}{c} = \begin{bmatrix} A_{1,1}/c & A_{1,2}/c & \cdots & A_{1,n}/c \\ A_{2,1}/c & A_{2,2}/c & \cdots & A_{2,n}/c \\ \vdots & \vdots & \ddots & \vdots \\ A_{m,1}/c & A_{1,1}/c & \cdots & A_{m,n}/c \end{bmatrix}$$

Basic Linear Algebra

Scalar Addition/Subtraction (Broadcasting)

$$c + \boldsymbol{A} = \begin{bmatrix} c + A_{1,1} & c + A_{1,2} & \cdots & c + A_{1,n} \\ c + A_{2,1} & c + A_{2,2} & \cdots & c + A_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ c + A_{m,1} & c + A_{1,1} & \cdots & c + A_{m,n} \end{bmatrix}$$

$$c - \boldsymbol{A} = \begin{bmatrix} c - A_{1,1} & c - A_{1,2} & \cdots & c - A_{1,n} \\ c - A_{2,1} & c - A_{2,2} & \cdots & c - A_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ c - A_{m,1} & c - A_{1,1} & \cdots & c - A_{m,n} \end{bmatrix}$$

Basic Linear Algebra

Broadcasting a Vector to a Matrix

$$\boldsymbol{v} + \boldsymbol{A} = \begin{bmatrix} v_1 + A_{1,1} & v_1 + A_{1,2} & \cdots & v_1 + A_{1,n} \\ v_2 + A_{2,1} & v_2 + A_{2,2} & \cdots & v_2 + A_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ v_m + A_{m,1} & v_m + A_{m,2} & \cdots & v_m + A_{m,n} \end{bmatrix}$$

where $\boldsymbol{v} \in \mathbb{R}^m$ and $\boldsymbol{A} \in \mathbb{R}^{m \times n}$

Basic Linear Algebra

Broadcasting a Vector to a Matrix

$$\boldsymbol{v}^\top + \boldsymbol{A} = \begin{bmatrix} v_1 + A_{1,1} & v_2 + A_{1,2} & \cdots & v_n + A_{1,n} \\ v_1 + A_{2,1} & v_2 + A_{2,2} & \cdots & v_n + A_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ v_1 + A_{m,1} & v_2 + A_{m,2} & \cdots & v_n + A_{m,n} \end{bmatrix}$$

where $\boldsymbol{v} \in \mathbb{R}^n$ and $\boldsymbol{A} \in \mathbb{R}^{m \times n}$

Basic Linear Algebra

Function Broadcasting

$$f(\boldsymbol{A}) = \begin{bmatrix} f(A_{1,1}) & f(A_{1,2}) & \cdots & f(A_{1,n}) \\ f(A_{2,1}) & f(A_{2,2}) & \cdots & f(A_{2,n}) \\ \vdots & \vdots & \ddots & \vdots \\ f(A_{m,1}) & f(A_{1,1}) & \cdots & f(A_{m,n}) \end{bmatrix}$$

where $f: \mathbb{R} \to \mathbb{R}$

Basic Linear Algebra

Dot Product

$$\boldsymbol{a}^\top \boldsymbol{b} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n = \sum_{i=1}^{n} a_i b_i$$

where $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$

Basic Linear Algebra

Matrix Multiplication

$$\boldsymbol{AB} = \begin{bmatrix} A_{1,:}B_{:,1} & A_{1,:}B_{:,2} & \cdots & A_{1,:}B_{:,n} \\ A_{2,:}B_{:,1} & A_{2,:}B_{:,2} & \cdots & A_{2,:}B_{:,n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{l,:}B_{:,1} & A_{l,:}B_{:,2} & \cdots & A_{l,:}B_{:,n} \end{bmatrix} \in \mathbb{R}^{l\times n}$$

where $\boldsymbol{A} \in \mathbb{R}^{l\times m}$ and $\boldsymbol{B} \in \mathbb{R}^{m\times n}$

Basic Linear Algebra

Transpose

$$\boldsymbol{A}^{\top} = \begin{bmatrix} A_{1,1} & A_{2,1} & \cdots & A_{m,1} \\ A_{1,2} & A_{2,2} & \cdots & A_{m,2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1,n} & A_{2,n} & \cdots & A_{m,n} \end{bmatrix} \in \mathbb{R}^{n \times m}$$

where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$

Basic Linear Algebra

Transpose

$$(\boldsymbol{A} + \boldsymbol{B})^{\top} = \boldsymbol{A}^{\top} + \boldsymbol{B}^{\top}$$

$$(\boldsymbol{A}\boldsymbol{B})^{\top} = \boldsymbol{B}^{\top}\boldsymbol{A}^{\top}$$

Basic Linear Algebra

Vector Norm

$$\|\boldsymbol{v}\| = \sqrt{\boldsymbol{v}^\top \boldsymbol{v}}$$

Basic Linear Algebra

Zero and One Vectors

$$\mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad \mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

Basic Linear Algebra

Linear Map

$$f(\boldsymbol{x}) = \boldsymbol{W}\boldsymbol{x} + \boldsymbol{b}$$

Bilinear Map

$$g(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^\top \boldsymbol{W}\boldsymbol{y} + \boldsymbol{b}$$

Basic Linear Algebra

Let $\mathbb{B} \subseteq \mathbb{R}^n$ be a set of vectors. The *span of* $\mathbb{B}$ is the set

$$\mathbf{span}(\mathbb{B}) = \{a_1 \boldsymbol{b}^{(1)} + a_2 \boldsymbol{b}^{(2)} + \cdots + a_k \boldsymbol{b}^{(k)} \mid \boldsymbol{a} \in \mathbb{R}^k, \forall i[\boldsymbol{b}^{(i)} \in \mathbb{B}]\}$$

A set of vectors $\mathbb{V}$ is a *vector space* if $\mathbb{V} = \mathrm{span}(\mathbb{B})$ for some $\mathbb{B}$.

Basic Linear Algebra

A set of vectors $\mathbb{B}$ is *linearly independent* if

$$a_1 \boldsymbol{b}^{(1)} + a_2 \boldsymbol{b}^{(2)} + \cdots + a_k \boldsymbol{b}^{(k)} = \mathbf{0}$$

implies $\boldsymbol{a} = \mathbf{0}$ for all $\boldsymbol{b}^{(1)}, \boldsymbol{b}^{(2)}, \ldots, \boldsymbol{b}^{(k)} \in \mathbb{B}$.

$\mathbb{B}$ is a *basis* for vector space $\mathbb{V}$ if $\mathbb{B}$ is linearly independent and $\operatorname{span}(\mathbb{B}) = \mathbb{V}$. The *dimension of* $\mathbb{V}$ is $\dim(\mathbb{V}) = |\mathbb{B}|$.

Feature Vectors

	Basil	Crushed Italian Tomatoes	Fresh Basil	Fresh Clams	Fresh Mozzarella	Fresh Native Tomatoes	Garlic	Grated Pecorino Romano	Meatball	Mozzarella	Olive Oil	Oregano	Ricotta Cheese
White Clam Pizza	0	0	0	1	0	0	1	1	0	0	1	1	0
Frank Pepe's Original Tomato Pie	0	1	0	0	0	0	0	1	0	0	1	0	0
Margherita	0	1	1	0	1	0	0	1	0	0	1	0	0
Fresh Tomato Pie	1	0	0	0	0	1	1	1	0	1	1	0	0
Meatball & Ricotta	0	1	0	0	0	0	0	1	1	1	1	0	1

What does $\boldsymbol{p} + \boldsymbol{q}$ represent?

Feature Vectors

	Basil	Crushed Italian Tomatoes	Fresh Basil	Fresh Clams	Fresh Mozzarella	Fresh Native Tomatoes	Garlic	Grated Pecorino Romano	Meatball	Mozzarella	Olive Oil	Oregano	Ricotta Cheese
White Clam Pizza	0	0	0	1	0	0	1	1	0	0	1	1	0
Frank Pepe's Original Tomato Pie	0	1	0	0	0	0	0	1	0	0	1	0	0
Margherita	0	1	1	0	1	0	0	1	0	0	1	0	0
Fresh Tomato Pie	1	0	0	0	0	1	1	1	0	1	1	0	0
Meatball & Ricotta	0	1	0	0	0	0	0	1	1	1	1	0	1

What does $\boldsymbol{p} - \boldsymbol{q}$ represent?

Feature Vectors

$$
\begin{array}{r|ccccccccccccc}
 & \text{Basil} & \text{Crushed Italian Tomatoes} & \text{Fresh Basil} & \text{Fresh Clams} & \text{Fresh Mozzarella} & \text{Fresh Native Tomatoes} & \text{Garlic} & \text{Grated Pecorino Romano} & \text{Meatball} & \text{Mozzarella} & \text{Olive Oil} & \text{Oregano} & \text{Ricotta Cheese} \\
\hline
\text{White Clam Pizza} & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\
\text{Frank Pepe's Original Tomato Pie} & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
\text{Margherita} & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
\text{Fresh Tomato Pie} & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\
\text{Meatball \& Ricotta} & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1
\end{array}
$$

What does $\boldsymbol{p} \odot \boldsymbol{q}$ represent?

Feature Vectors

	Basil	Crushed Italian Tomatoes	Fresh Basil	Fresh Clams	Fresh Mozzarella	Fresh Native Tomatoes	Garlic	Grated Pecorino Romano	Meatball	Mozzarella	Olive Oil	Oregano	Ricotta Cheese
White Clam Pizza	0	0	0	1	0	0	1	1	0	0	1	1	0
Frank Pepe's Original Tomato Pie	0	1	0	0	0	0	0	1	0	0	1	0	0
Margherita	0	1	1	0	1	0	0	1	0	0	1	0	0
Fresh Tomato Pie	1	0	0	0	0	1	1	1	0	1	1	0	0
Meatball & Ricotta	0	1	0	0	0	0	0	1	1	1	1	0	1

What does $\max(\boldsymbol{p}, \boldsymbol{q})$ represent?

Feature Vectors

	Basil	Crushed Italian Tomatoes	Fresh Basil	Fresh Clams	Fresh Mozzarella	Fresh Native Tomatoes	Garlic	Grated Pecorino Romano	Meatball	Mozzarella	Olive Oil	Oregano	Ricotta Cheese
White Clam Pizza	0	0	0	1	0	0	1	1	0	0	1	1	0
Frank Pepe's Original Tomato Pie	0	1	0	0	0	0	0	1	0	0	1	0	0
Margherita	0	1	1	0	1	0	0	1	0	0	1	0	0
Fresh Tomato Pie	1	0	0	0	0	1	1	1	0	1	1	0	0
Meatball & Ricotta	0	1	0	0	0	0	0	1	1	1	1	0	1

What does $\boldsymbol{p}^\top \boldsymbol{q}$ represent?

Feature Vectors

- $\boldsymbol{p}$ = white pie with meatballs
- $\boldsymbol{q}$ = plain white pie
- $\boldsymbol{r}$ = red pie with meatballs
- $\boldsymbol{s}$ = plain red pie
- What does $\boldsymbol{p} - \boldsymbol{q} + \boldsymbol{s}$ represent?

Similarity and Distance Metrics

Cosine Similarity

$$\cos(\boldsymbol{a}, \boldsymbol{b}) = \frac{\boldsymbol{a}^{\top}\boldsymbol{b}}{\|\boldsymbol{a}\|\,\|\boldsymbol{b}\|}$$

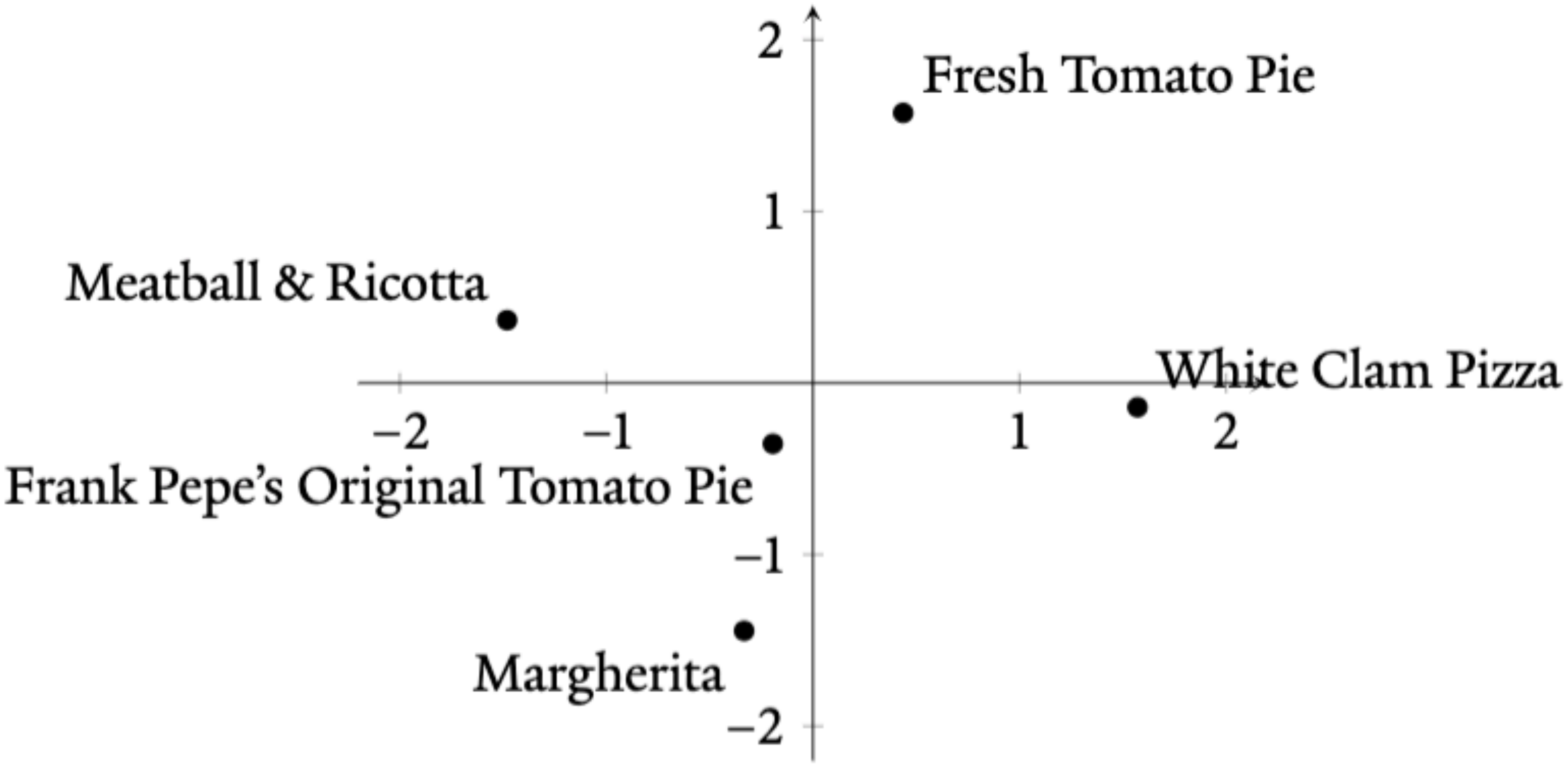
Cosine Distance

$$1 - \cos(\boldsymbol{a}, \boldsymbol{b})$$

Euclidean (l^2) Distance

$$\|\boldsymbol{a} - \boldsymbol{b}\|$$

Embeddings



This embedding is roughly isometric.

Embeddings

- An *embedding of* $\mathbb{A}$ *into* $\mathbb{B}$ is a one-to-one mapping $f: \mathbb{A} \to \mathbb{B}$.
- f is *isometric with respect to* d_1 *and* d_2 if for all $x, y \in \mathbb{A}$, $d_1(x, y) = d_2(f(x), f(y))$.
- If $\dim(\mathbb{B}) < \dim(\mathbb{A})$, then f is performing *dimensionality reduction*.

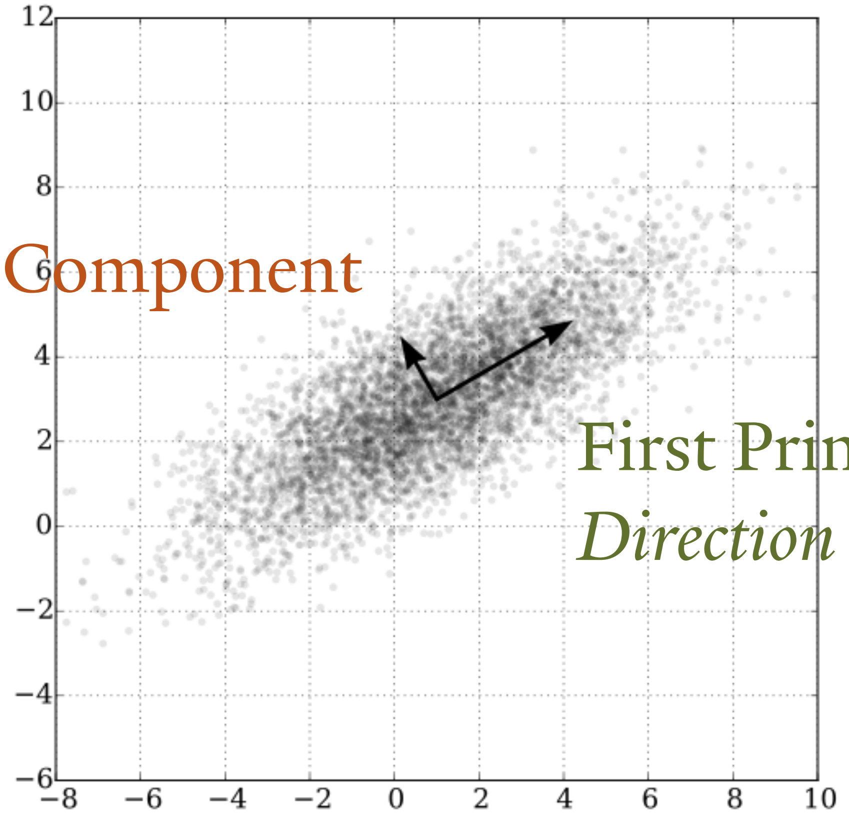
Algorithms for Dimensionality Reduction

- Principal Components Analysis (PCA)
- t-Distributed Stochastic Neighbor Embedding (t-SNE)

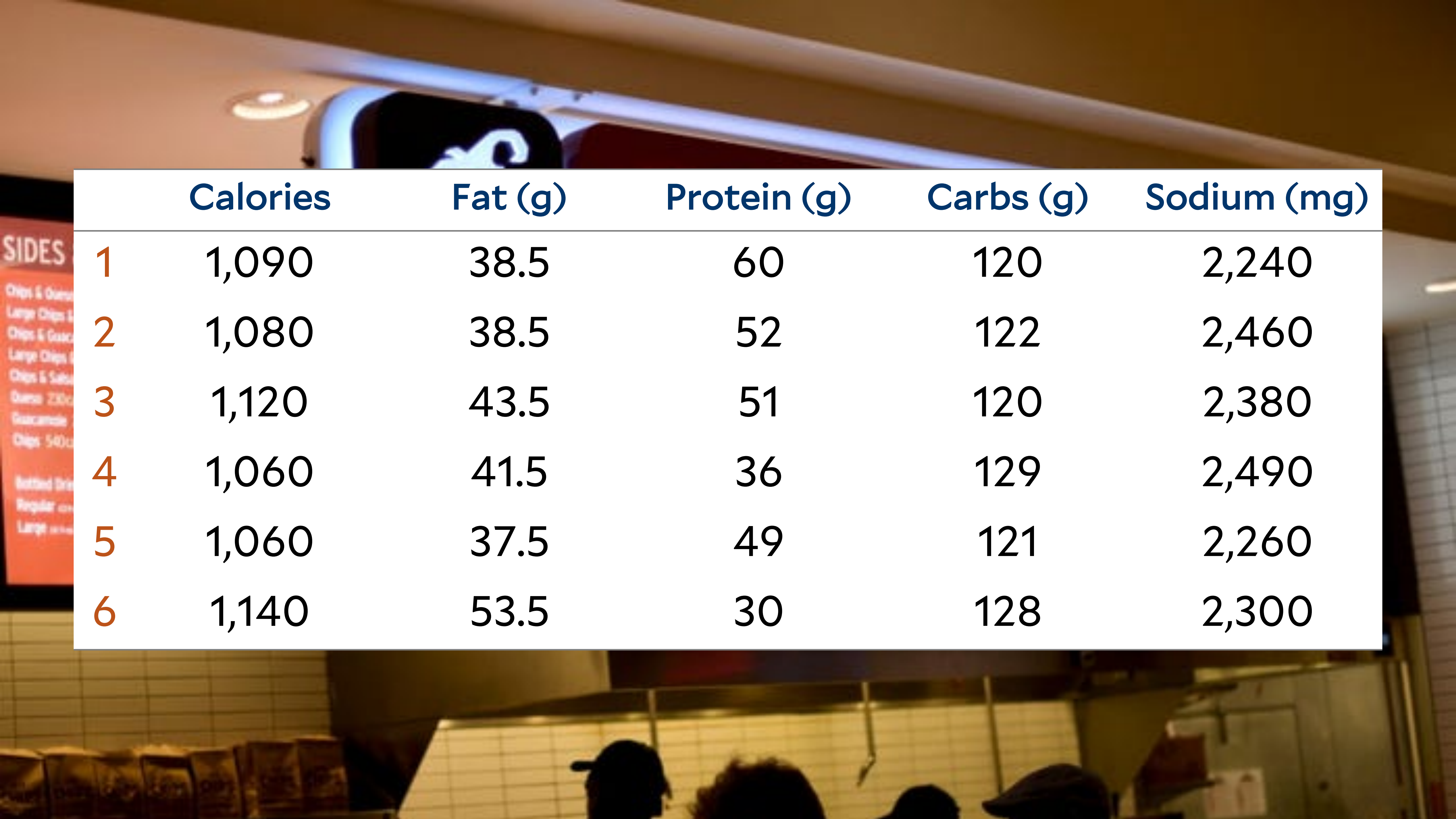
Learn more in Course Notes Section 1.5!

Principal Components Analysis

Second Principal Component



First Principal Component
Direction of Greatest Variance



	Calories	Fat (g)	Protein (g)	Carbs (g)	Sodium (mg)
1	1,090	38.5	60	120	2,240
2	1,080	38.5	52	122	2,460
3	1,120	43.5	51	120	2,380
4	1,060	41.5	36	129	2,490
5	1,060	37.5	49	121	2,260
6	1,140	53.5	30	128	2,300

	Tortilla	Barbacoa	Carnitas	Chicken	Sofritas	Steak	Brown Rice	White Rice	Black Beans	Pinto Beans	Tomato Salsa	Guacamole	Cheese	Sour Cream	Green Salsa	Red Salsa	Lettuce
1	1	0	0	1	0	0	0	1	1	0	1	0	1	1	0	0	1
2	1	1	0	0	0	0	0	1	1	0	1	0	1	1	0	0	1
3	1	0	1	0	0	0	0	1	1	0	1	0	1	1	0	0	1
4	1	0	0	0	1	0	0	1	1	0	1	0	1	1	0	0	1
5	1	0	0	0	0	1	0	1	1	0	1	0	1	1	0	0	1
6	1	0	0	0	0	0	0	1	1	0	1	1	1	1	0	0	1

	Chicken Burrito	Barbacoa Burrito	Carnitas Burrito	Sofritas Burrito	Steak Burrito	Veggie Burrito
1	1	0	0	0	0	0
2	0	1	0	0	0	0
3	0	0	1	0	0	0
4	0	0	0	1	0	0
5	0	0	0	0	1	0
6	0	0	0	0	0	1

	Calories	Fat	Protein	Carbs	Sodium
1	55%	49%	120%	44%	97%
2	54%	49%	104%	44%	107%
3	56%	56%	102%	44%	103%
4	53%	53%	72%	47%	108%
5	53%	48%	98%	44%	98%
6	57%	69%	60%	47%	100%

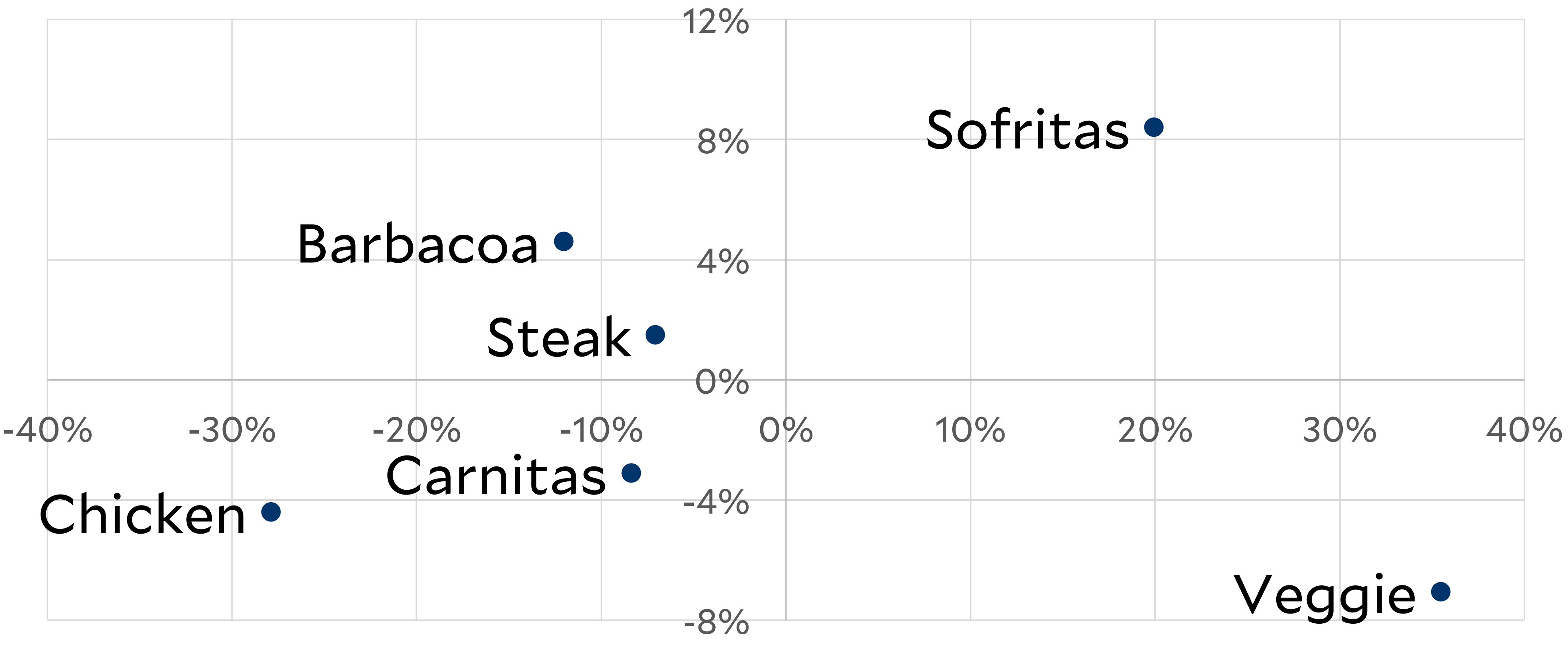
Variance of Features

	Calories	Fat	Protein	Carbs	Sodium
1	55%	49%	120%	44%	97%
2	54%	49%	104%	44%	107%
3	56%	56%	102%	44%	103%
4	53%	53%	72%	47%	108%
5	53%	48%	98%	44%	98%
6	57%	69%	60%	47%	100%
σ^2	0.026%	0.590%	4.971%	0.022%	0.209%

Latent Features (Principal Components)

	PC1	PC2	PC3	PC4	PC5
1	−27.90%	−4.39%	1.33%	0.81%	0.00%
2	−12.03%	4.62%	−3.45%	−0.34%	0.00%
3	−8.39%	−3.09%	−3.81%	−0.37%	0.00%
4	19.94%	8.42%	0.18%	0.57%	0.00%
5	−7.07%	1.50%	5.74%	−0.66%	0.00%
6	35.45%	−7.05%	0.01%	−0.02%	0.00%
σ^2	5.395%	0.346%	0.122%	0.003%	0.00%

Visualizing Principal Components



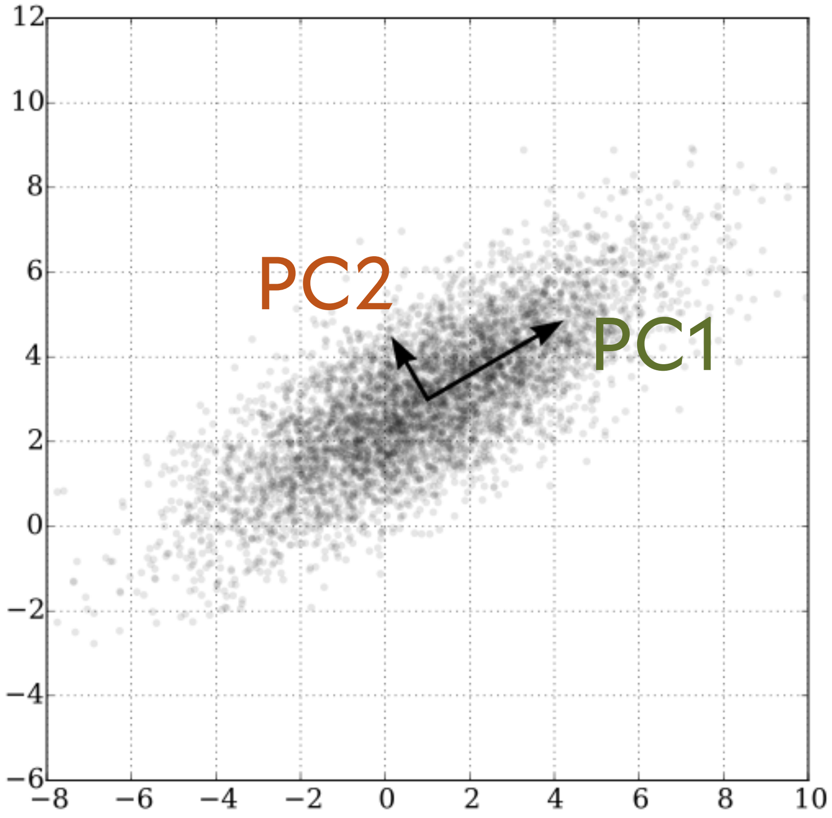
Latent Features (Principal Components)

Principal Components

$$\boldsymbol{\Psi} = \begin{bmatrix} 1.81\% & 27.32\% & -95.82\% & 6.16\% & 5.48\% \\ -24.07\% & -75.46\% & -18.49\% & 2.36\% & 58.13\% \\ -19.90\% & -51.33\% & -19.60\% & 0.96\% & -81.14\% \\ 5.68\% & -1.02\% & 6.16\% & 99.64\% & -1.06\% \\ 94.81\% & -30.39\% & -7.35\% & -5.29\% & -2.31\% \end{bmatrix}$$

where $(\boldsymbol{\Psi}_{1,:})^{\top}$ = PC1, $(\boldsymbol{\Psi}_{2,:})^{\top}$ = PC2, etc.

Principal Components Analysis

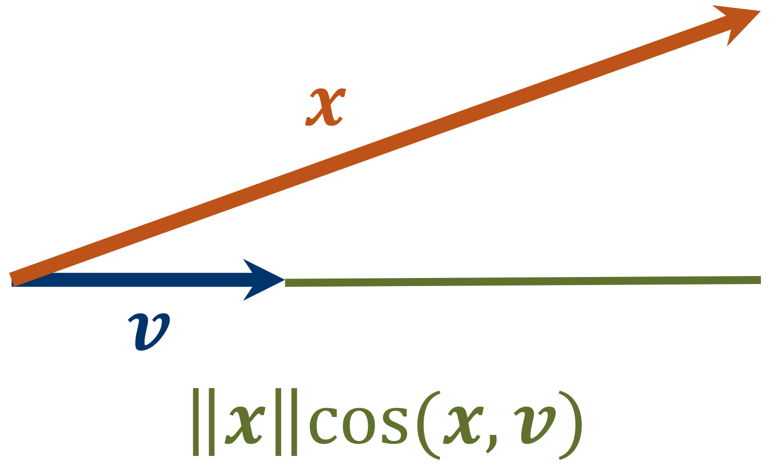


Principal Components Analysis

PC1:

$$\boldsymbol{\psi}^{(1)} = \underset{\boldsymbol{v}}{\operatorname{argmax}} \operatorname{Var}(\{\|\boldsymbol{x}\|\cos(\boldsymbol{x}, \boldsymbol{v}) \mid \boldsymbol{x} \in \mathbb{D}\})$$

$$\text{subject to } \|\boldsymbol{v}\| = 1$$



Principal Components Analysis

PC1:

$$\boldsymbol{\psi}^{(1)} = \underset{\boldsymbol{v}}{\operatorname{argmax}} \operatorname{Var}(\{\|\boldsymbol{x}\|\cos(\boldsymbol{x},\boldsymbol{v}) \mid \boldsymbol{x} \in \mathbb{D}\})$$

$$\text{subject to } \|\boldsymbol{v}\| = 1$$

$$\|\boldsymbol{x}\|\cos(\boldsymbol{x},\boldsymbol{v}) = \|\boldsymbol{x}\| \frac{\boldsymbol{x}^\top \boldsymbol{v}}{\|\boldsymbol{x}\|\|\boldsymbol{v}\|} = \boldsymbol{x}^\top \boldsymbol{v}$$

Principal Components Analysis

PC1:

$$\boldsymbol{\psi}^{(1)} = \underset{\boldsymbol{v}}{\operatorname{argmax}} \operatorname{Var}(\{\boldsymbol{x}^\top \boldsymbol{v} \mid \boldsymbol{x} \in \mathbb{D}\})$$
$$\text{subject to } \|\boldsymbol{v}\| = 1$$

PC2, PC3, etc.:

$$\boldsymbol{\psi}^{(i)} = \underset{\boldsymbol{v}}{\operatorname{argmax}} \operatorname{Var}(\{\boldsymbol{x}^\top \boldsymbol{v} \mid \boldsymbol{x} \in \mathbb{D}\})$$
$$\text{subject to } \|\boldsymbol{v}\| = 1 \text{ and } \boldsymbol{v}^\top \boldsymbol{\psi}^{(j)} = 0 \text{ for all } j < i$$