# Linear Classifiers (Part 1)

CS114B Lab 2

Kenneth Lai

February 3, 2022

# Vectors and Vector Spaces

- Wikipedia: A vector space is a set of objects called vectors, which may be added together and multiplied by numbers called scalars

# Vectors and Vector Spaces

- Wikipedia: A vector space is a set of objects called vectors, which may be added together and multiplied by numbers called scalars
- Lots of examples of vector spaces; we will focus on a particular type of real vector space/coordinate space, where
  - Vectors are $n$-tuples of real numbers, for some natural number $n$
  - Scalars are real numbers

# Vectors and Vector Spaces

- Wikipedia: A vector space is a set of objects called vectors, which may be added together and multiplied by numbers called scalars
- Lots of examples of vector spaces; we will focus on a particular type of real vector space/coordinate space, where
  - Vectors are elements of $\mathbb{R}^n$
  - Scalars are elements of $\mathbb{R}$

# Vectors and Vector Spaces

- Wikipedia: A vector space is a set of objects called vectors, which may be added together and multiplied by numbers called scalars
- Lots of examples of vector spaces; we will focus on a particular type of real vector space/coordinate space, where
  - Vectors are elements of $\mathbb{R}^n$
  - Scalars are elements of $\mathbb{R}$
- The dimension of such a vector space (not to be confused with a dimension, i.e., axis, of a Numpy array) is $n$

# Vectors in NLP

- Idea: objects of interest (e.g., documents, words, etc.) can be represented as vectors (embeddings) in some vector space

# Vectors in NLP

- Idea: objects of interest (e.g., documents, words, etc.) can be represented as vectors (embeddings) in some vector space
- Coordinates of the vector correspond to features of the object

# Vectors in NLP

- Idea: objects of interest (e.g., documents, words, etc.) can be represented as vectors (embeddings) in some vector space
- Coordinates of the vector correspond to features of the object
  - Sometimes, these features are human-interpretable
    - Naïve Bayes features: word counts in a document

# Vectors in NLP

- Idea: objects of interest (e.g., documents, words, etc.) can be represented as vectors (embeddings) in some vector space
- Coordinates of the vector correspond to features of the object
  - Sometimes, these features are human-interpretable
    - Naïve Bayes features: word counts in a document
  - Sometimes, they are not
    - Many word vector "features"

# Linear Classifiers

- Suppose we have a classification problem (e.g., text classification, sequence labeling, etc.)

# Linear Classifiers

- Suppose we have a classification problem (e.g., text classification, sequence labeling, etc.)
- Linear classifiers make their classification decisions based on a linear combination of features

# Linear Classifiers

- Suppose we have a classification problem (e.g., text classification, sequence labeling, etc.)
- Linear classifiers make their classification decisions based on a linear combination of features
  - Logistic regression
  - Perceptron
  - Naïve Bayes (in a way)
  - ...

# Linear Classifiers

- Let $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}$ be a feature vector, $\theta = \begin{bmatrix} \theta_1 & \dots & \theta_n \end{bmatrix}$ be a vector of parameters, $g$ be some function, $\hat{y}$ be the classification decision, and $\cdot$ denote the dot product

# Linear Classifiers

- Let $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}$ be a feature vector, $\theta = \begin{bmatrix} \theta_1 & \dots & \theta_n \end{bmatrix}$ be a vector of parameters, $g$ be some function, $\hat{y}$ be the classification decision, and $\cdot$ denote the dot product
$$\hat{y} = g\left( \sum_{j=1}^{n} \theta_j x_j \right) = g(\theta \cdot \mathbf{x})$$

# Linear Classifiers

- Sometimes (especially in logistic regression), within the set of parameters, we distinguish between weights $w_j$ and a bias term $b$

# Linear Classifiers

- Sometimes (especially in logistic regression), within the set of parameters, we distinguish between weights $w_j$ and a bias term $b$
  - This is equivalent to having a "dummy feature" with value 1

# Linear Classifiers

- Sometimes (especially in logistic regression), within the set of parameters, we distinguish between weights $w_j$ and a bias term $b$
  - This is equivalent to having a "dummy feature" with value 1

$$\theta = \begin{bmatrix} w_1 & \dots & w_n & b \end{bmatrix} = \left[ \begin{array}{c|c} \mathbf{w} & b \end{array} \right]$$

$$\mathbf{x}' = \begin{bmatrix} x_1 & \dots & x_n & 1 \end{bmatrix} = \left[ \begin{array}{c|c} \mathbf{x} & 1 \end{array} \right]$$

$$\hat{y} = g\left( \sum_{j=1}^{n} w_j x_j + b \right) = g(\mathbf{w} \cdot \mathbf{x} + b) = g(\theta \cdot \mathbf{x}')$$