CSCE 790: Neural Networks and Their Applications

AIISC and Dept. Computer Science and Engineering
Email: vignar@sc.edu

Dr. Vignesh Narayanan

September 5, 2023

UNIVERSITY OF
South Carolina

Projects - Start Early!

Your overall final course letter grade will be determined by your grades on the following assessments.

Homework Assignment	15%
Presentation	15%
Midterm Exam (Take home)	15%
Final Project	55%

Supervised Learning

- Input and target outputs are given for training
- Learning relationship between the input output pairs
- Types:
  - **Regression:** Covers situations where Y is continuous (quantitative)
  - Example: predicting the value of the Dow in 6 months, predicting the value of a given house based on various inputs, etc.
  - **Classification:** Covers situations where Y is categorical (qualitative)
  - Example: Will the Dow be up or down in 6 months? Is this email spam or not?

Potential Paper Presentation or Project Topic - 1 (Image analysis)

Decision Boundaries

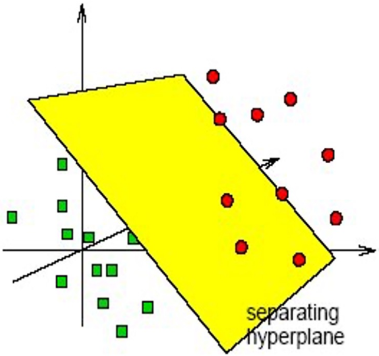

Figure: Linear DB

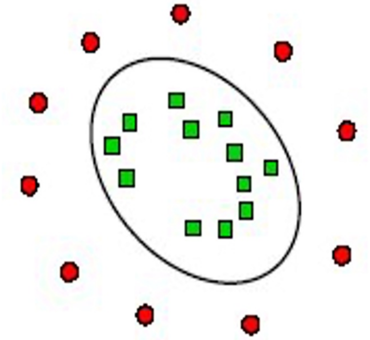

Figure: Nonlinear DB

Second Detour - Linear Algebra - Spaces

Definition (Vector Space)

A vector space V over a field $\mathbb{F}$ is a set of elements called vectors, together with two operations, addition: $V \times V \to V$, $x, y \in V \mapsto x + y \in V$, and scalar multiplication: $\mathbb{F} \times V \to V$, $\alpha \in \mathbb{F}$, $x \in V \mapsto \alpha x \in V$, satisfying 8 properties (refer slides from last class).

Vector Space

Example (Vector spaces)

(i) $\{0\}$, the trivial space.

(ii) $\mathbb{R}$ over $\mathbb{R}$.

(iii) $\mathbb{R}^n$ over $\mathbb{R}$.

(iv) Space of $m \times n$ matrices, $\mathbb{R}^{m\times n}$ over $\mathbb{R}$.

(v) The collection of all real-valued continuous functions $f : [a, b] \to \mathbb{R}$ over $[a, b] \subset \mathbb{R}$, denoted as $C[a, b]$ with $\mathbb{F} = \mathbb{R}$. $x = y$ if $x(t) = y(t)$, $\forall t \in [a, b]$.

(vi) $C^m[t_0, t_1]$, the space of m-tuples $f : [t_0, t_1] \to \mathbb{R}^m$ whose elements are continuous functions on $[t_0, t_1]$.

Subspace

Definition (Subspace)

A nonempty subset S of a vector space V is called a subspace of V if $\alpha x + \beta y \in S$ for every $x, y \in S$ and every $\alpha, \beta \in \mathbb{R}$.

Remarks

1. By definition, a subspace must contain the null vector 0.
2. V is itself a subspace of V.
3. A subspace not equal to the entire space is said to be a proper subspace.

Potential Paper Presentation or Project Topic - 2 (NLP)

Span

Definition (Span)

Let V be a vector space. Given $x_1, \ldots, x_m \in V$, the span of $x_1, \ldots, x_m$, denoted by $\mathrm{span}\{x_1, \ldots, x_m\}$, is the set of all vectors v that can be written as $v = \sum_{i=1}^{m} \alpha_i x_i$ for some $\alpha_i \in \mathbb{R}$. That is,

$$\mathrm{span}\{x_1, \ldots, x_m\} = \{v \in V : v = \sum_{i=1}^{m} \alpha_i x_i \text{ for some} \alpha_i \in \mathbb{R}\}.$$

We say v can be written as a linear combination of the vectors $x_1, \ldots, x_m$.

Linear Independence

Definition (Linear Independence)

A set of vectors $x_1, \ldots, x_k$ in a vector space V is said to be linearly independent if $\sum_{i=1}^{m} \alpha_i x_i = 0$, where $\alpha_1, \ldots, \alpha_m$ are constants, implies that $\alpha_i = 0$ for all $i = 1, \ldots, m$. That is,

$$\sum_{i=1}^{m} \alpha_i x_i = 0 \implies \alpha_i = 0, \forall i = 1, \ldots m.$$

Definition (Basis)

If $\mathrm{span}\{x_1, \ldots, x_n\} = V$ and $\{x_1, \ldots, x_n\}$ is a linearly independent set, it is said to be a basis of V.

Inner Product

Definition (Inner Product)

Let X be a real vector space. An inner product on X is a mapping $\langle \cdot, \cdot \rangle : X \times X \to \mathbb{R}$ such that $\forall x, y, z \in X, a, b \in \mathbb{R}$, we have

1. $\langle x, x \rangle \geq 0$ (positivity) and $\langle x, x \rangle = 0 \iff x = 0$ (definiteness)
2. $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$ (additivity in the first slot)
3. $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$ (homogeneity in the first slot)
4. $\langle x, y \rangle = \overline{\langle y, x \rangle}$ (conjugate symmetry)

Two vectors $x, y \in V$ are said to be orthogonal if $\langle x, y \rangle = 0$.

Norm

Definition (Norm)

A mapping $\|\cdot\| : X \to \mathbb{F}$ on a vector space X over a field $\mathbb{F}$ is called a norm if for all $x, y \in X$ and $\alpha \in \mathbb{F}$ it satisfies

(i) $\|x\| \geq 0$ (positivity) and $\|x\| = 0 \iff x = 0$ (definiteness)

(ii) $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality)

(iii) $\|\alpha x\| = |\alpha| \cdot \|x\|$ (homogeneity).

If X is an inner product space, then $\|x\| = \sqrt{\langle x, x \rangle}$ is the induced norm.

Example

Example

1. The Euclidean norm (or ℓ_2 norm) on $\mathbb{R}^n$ is $\|x\|_2 = \sqrt{\langle x, x \rangle} = \sum_{i=1}^{n} x_i^{2\,\frac{1}{2}}$.
2. The Supremum norm (or ℓ_∞ norm) on $\mathbb{R}^n$ is $\|x\|_\infty = \max\{|x_1|, \ldots, |x_n|\}$.
3. The matrix norm (or induced norm or operator norm) of an $n \times n$ matrix A is

$$\|A\| = \max_{\{x \in \mathbb{R}^n \,|\, \|x\| = 1\}} \|Ax\| = \max_{x \in \mathbb{R}^n} \frac{\|Ax\|}{\|x\|}.$$

   Note that there are various matrix norms, here we are interested in the *induced norm* or *operator norm*.

4. The p-norm of $\mathbb{R}^n$ is $\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}$ for $p \geq 1$.

Equivalence of Norms

Proposition (Equivalence of norms on $\mathbb{R}^n$)

For any two norms $\|\cdot\|_a$ and $\|\cdot\|_b$ on $\mathbb{R}^n$, there exist positive constants α and β such that for all $x \in \mathbb{R}^n$,

$$\alpha\|x\|_a \leq \|x\|_b \leq \beta\|x\|_a.$$

Example

$\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n}\|x\|_\infty$. For example, if $x = (3,4)^T$, then $\|x\|_\infty = 4$, $\|x\|_2 = 5$ and hence $\sqrt{n}\|x\|_\infty = 4\sqrt{2}$.

Square Matrices and Eigenvalues

Definition (Singular matrix)

An $n \times n$ matrix A is called singular if $\det(A) = 0$. It is nonsingular or invertible if $\det(A) \neq 0$. If A and B are $n \times n$ invertible matrices, then $(AB)^{-1} = B^{-1}A^{-1}$.

Proposition

Let A be a square matrix.

1. *A complex number λ is an eigenvalue of A if and only if there exists a nonzero eigenvector associated with λ.*
2. *A is singular if and only if it has an eigenvalue that is equal to zero.*

Square Matrices and Eigenvalues

Definition (Spectral Radius)

The spectral radius of a matrix $A \in \mathbb{R}^{n \times n}$ is $\rho(A) = \max|\lambda_i|$, where $\lambda(A) = \{\lambda_1, \ldots, \lambda_n\}$, namely, the maximum of the magnitudes of the eigenvalues of A. Then $\rho : \mathbb{R}^{n \times n} \to \mathbb{R}$ is a constant function of A.

Revisiting Parametric Models

- Given data:

$$\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$$

- Model Choice:

$$\hat{y}_i = f_\theta(x_i) = \theta_0 + \theta_1 x_{i1} + \theta_2 x_{i2} + \ldots + \theta_n x_{ip}$$

- Let $x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix}$

- Let $\theta = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_p \end{pmatrix}$

- n, p are number of samples and number of features per sample
- $f_\theta(x)$ is a linear model

Cost Function

- We can measure the accuracy of our hypothesis function by using a cost function

$$J(\theta) = \frac{1}{n} \sum_{i=1,\ldots,n} (\hat{y}_i - y_i)^2 = \frac{1}{n} \sum_{i=1,\ldots,n} (f_\theta(x_i) - y_i)^2$$

- Find θ such that the predicted output is close to the actual output

$$\min_{\theta \in \mathbb{R}^p} J(\theta)$$

Example 2-Parameter Model

- For a fixed θ, $f_\theta(x)$ is a function of x
- Example:

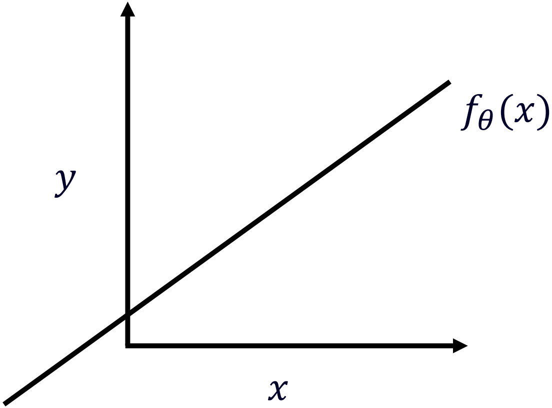

Figure: Example linear function for a fixed θ_0, θ_1

Cost Function

- The cost/objective/loss function is supported on the parameter space
- Example

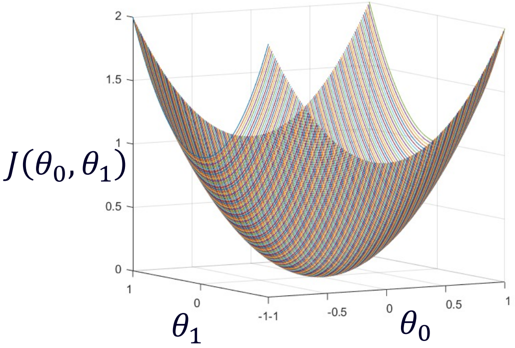

Figure: Example cost function supported in the two-dimensional parameter space (with θ_0, θ_1)

Third Detour - Optimization - Overview

Optimization problems are ubiquitous in science and engineering, and in our daily life. Thinking about how we

- optimize our way to go to work,
- choose the line we stand in at a supermarket, or
- maximize our profit in the stock market.

we are confronted in our daily life with making "optimal" decisions.

Third Detour - Optimization - Overview

Usually, we cannot freely choose from all available decision alternatives, but there are *constraints* that restrict the number of available alternatives. Common restrictions come from the following

- availability of resource
- law
- technical limitations
- interpersonal relations between humans

Third Detour - Optimization - Overview

Optimization models attempt to express, in mathematical terms, the goal of solving a problem in the "best" way. For example,

- running a business to maximize profit, minimize loss, maximize efficiency, or minimize risk,
- selecting a flight plan for an aircraft to minimize time or fuel use.

With the help of computer hardware and software, it is now possible to solve optimization problems with *thousands or even millions* of variables and constraints.

History of Optimization

- Fermat, 1638; Newton 1670:

$$\min \quad f(\theta)$$
$$\frac{df(\theta)}{d\theta} = 0.$$

- Euler 1755:

$$\min \quad f(\theta_1, \ldots, \theta_n)$$
$$\nabla f(\theta) = 0.$$

History of Optimization

- Lagrange, 1797:

$$\begin{aligned} \text{min} \quad & f(\theta_1, \ldots, \theta_n) \\ \text{s.t.} \quad & g_k(\theta_1, \ldots, \theta_n) = 0, \quad k = 1, \ldots, m. \end{aligned}$$

- Euler, Lagrange: Problems in infinite dimensions, calculus of variations.