CSCE 790: Neural Networks and Their Applications

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

© Vignesh Narayanan

September 7, 2023

UNIVERSITY OF
South Carolina

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:

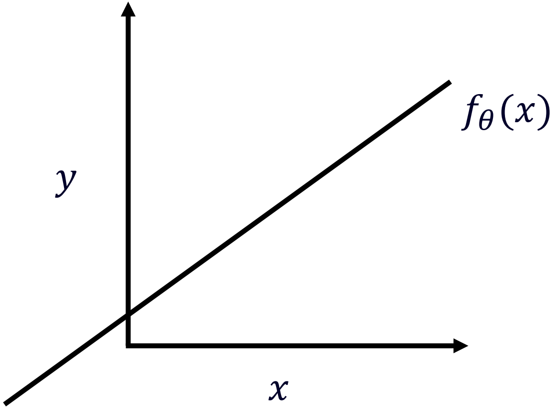

Figure: Example linear function for a fixed θ_0, θ_1

Cost Function

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

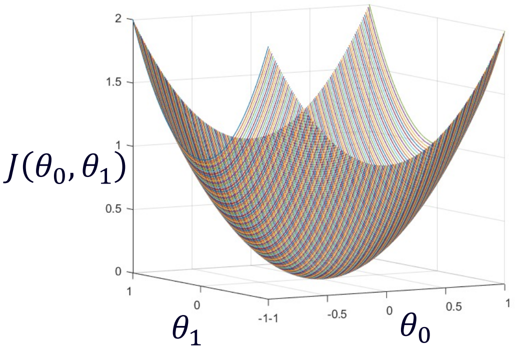

Figure: Example cost function supported in the two-dimensional parameter space (with θ_0, θ_1)

Mathematical Formulation

Mathematical models of optimization can be generally represented by

- f : a cost function (objective function)
- θ : available decisions (decision variables)
- Θ : a constraint set (feasible solutions),

where $f : \Theta \to \mathbb{R}$ and $\theta \in \Theta \subset \mathbb{R}^n$.

Definition (minimization problem)

Find an optimal decision, i.e., $\theta^* \in \Theta$, such that $f(\theta^*) \leq f(\theta), \forall \theta \in \Theta$.

Mathematical Formulation

Finite-dimensional problems, $\Theta \subseteq \mathbb{R}^n$.

- If $\Theta = \mathbb{R}^n$, then it is *unconstrained optimization*, i.e.,

$$\min_{\theta \in \mathbb{R}^n} \quad f(\theta).$$

- If $\Theta \subset \mathbb{R}^n$, then it is *constrained optimization*, i.e.,

$$\begin{aligned} \min_{\theta} \quad & f(\theta) \\ \text{s.t.} \quad & \theta \in \Theta \subset \mathbb{R}^n \end{aligned}$$

Types of Optimization

Linear Optimization

- The constraints and the objective function f are linear functions of the decision variables θ, namely, Θ is a polyhedron specified by linear inequality constraints.

Nonlinear Optimization

- The objective function or some or all of the constraints are represented with nonlinear functions.

Linear vs Nonlinear Function

Definition (Linear Function)

Let X and Y be vector spaces over the same field F. A function $f : X \to Y$ is called a linear map if for any two vectors $x_1, x_2 \in X$ and any scalar $a \in F$, the following conditions hold:

- (Superposition principle/Additivity) $f(x_1 + x_2) = f(x_1) + f(x_2)$
- (Homogeneity) $f(ax_1) = af(x_1)$

Definition (Nonlinear Function)

A function is nonlinear if it does not satisfy superposition or homogeneity.

Example (Linear Programming)

Solve the following minimization problem:

$$\begin{array}{llllll} \min_{x_1,x_2} & f = & -2x_1 & - & x_2 & \\ & x_1 & + & \frac{8}{3}x_2 & \leq 4 & \\ & x_1 & + & x_2 & \leq 2 & \\ & 2x_1 & & & \leq 3 & \\ & x_1, x_2 & \geq 0 & & & \end{array}$$

Example

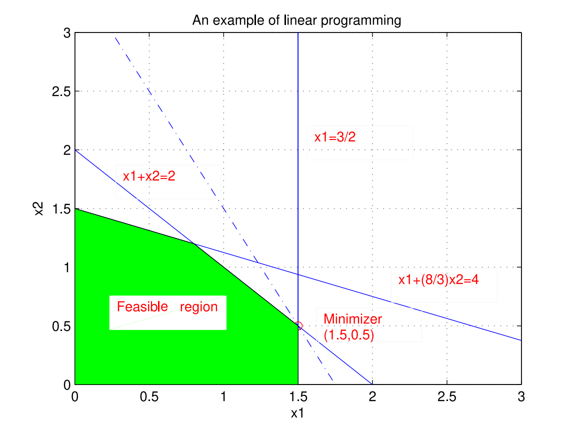

Figure: Illustration of the feasible region

Functions of One Variable

The fundamental results of calculus related to optimization are based on **Taylor's Formula** (or called the Extended Law of the Mean) for real-valued functions.

Theorem (Taylor's Formula; Extended Law of the Mean)

Suppose that $f(\theta)$, $f'(\theta)$, $f''(\theta)$ exist on the closed interval $[a, b]$. If θ^*, θ are any two different points of $[a, b]$, then there exists a point z strictly between θ^* and θ such that

$$f(\theta) = f(\theta^*) + f'(\theta^*)(\theta - \theta^*) + \frac{f''(z)}{2}(\theta - \theta^*)^2.$$

Relevance of Taylor's Forumla to Optimization

If $f''(\theta) > 0,\ \forall\,\theta$, and $f'(\theta^*) = 0$,

$$\begin{aligned}
&\implies f(\theta) = f(\theta^*) + 0 + \text{a positive number} \quad \forall\ \theta \neq \theta^* \\
&\implies f(\theta) > f(\theta^*) \quad \forall\ \theta \neq \theta^* \\
&\implies \theta^* \text{ is the minimizer of } f(\theta)
\end{aligned}$$

- Same reasoning that $f''(\theta) < 0$ and $f'(\theta^*) = 0$ are for maximizer.
- This is called the *Second Derivative Test*, which forms the basis of unconstrained optimization (via calculus).

Example

$$
\begin{aligned}
f(\theta) &= \exp^{\theta^2}, \\
f'(\theta) &= 2\theta \exp^{\theta^2}, \\
f''(\theta) &= 4\theta^2 \exp^{\theta^2} + 2\exp^{\theta^2} = (4\theta^2 + 2)\exp^{\theta^2} > 0 \quad \forall \theta \in \mathbb{R}.
\end{aligned}
$$

Since $f''(\theta) > 0$ for all real θ and since $f'(0) = 0$, we learn that $f(0) = 1$ is smaller than any other value of $f(\theta)$.

Functions of One Variable

Definition (Minimizers)

Suppose $f(\theta)$ is a real-valued function defined on some interval I (may be finite or infinite, open or closed, or half-open). A point $\theta^* \in I$ is

1. a global minimizer for $f(\theta)$ on I if $f(\theta^*) \leq f(\theta)$, $\forall\ \theta \in I$.
2. a strict global minimizer for $f(\theta)$ on I if $f(\theta^*) < f(\theta)$, $\forall\ \theta \in I$, such that $\theta \neq \theta^*$.
3. a local minimizer for $f(\theta)$ if there is a positive number δ such that $f(\theta^*) \leq f(\theta)$, $\forall\ \theta \in I$, for which $\theta^* - \delta < \theta < \theta^* + \delta$.
4. a strict local minimizer for $f(\theta)$ if there is a positive number δ such that $f(\theta^*) < f(\theta)$, $\forall\ \theta \in I$, for which $\theta^* - \delta < \theta < \theta^* + \delta$, $\theta \neq \theta^*$.
5. a **critical point** of $f(\theta)$ if $f'(\theta^*)$ exists and is equal to zero.

Functions of One Variable

Theorem

Suppose that $f(\theta)$ is differentiable on I. If θ^* is a local minimizer or maximizer of f, then either θ^* is an endpoint of I or $f'(\theta^*) = 0$.

Theorem

Suppose f, f', and f'' are all continuous on I and that $\theta^* \in I$ is a critical point of f.

a) *If $f''(\theta) \geq 0 \ \forall \theta \in I$, then θ^* is a global minimizer of $f(\theta)$ on I.*

b) *If $f''(\theta) > 0 \ \forall \theta \in I$ such that $\theta \neq \theta^*$, then θ^* is a strict global minimizer of $f(\theta)$ on I.*

c) *If $f''(\theta^*) > 0$, then θ^* is a strict local minimizer of $f(\theta)$.*

Functions of One Variable

Once the critical points of f have been identified, the previous result can be used to determine whether these points are minimizers. To test for maximizers, replace $f''(\theta) \geq 0$, $f''(\theta) > 0$, and $f''(\theta^*) > 0$ by $f''(\theta) \leq 0$, $f''(\theta) < 0$, and $f''(\theta^*) < 0$, respectively.

Functions of One Variable

Example

Find the minima of

$$f(\theta) = 3\theta^4 - 4\theta^3 + 1.$$

Here $f'(\theta) = 12\theta^3 - 12\theta^2 = 12\theta^2(\theta - 1)$, so the critical points are $\theta = 0$ and $\theta = 1$. $f''(\theta) = 36\theta^2 - 24\theta = 12\theta(3\theta - 2)$, so $f''(0) = 0$ and $f''(1) = 12$, so $\theta = 1$ is a strict local minimizer (by (c) of theorem stated above). But the theorem provides no information about $\theta = 0$. Note that because

(i) $\theta^4 < \theta^3$ for $0 < \theta < 1$ then $f(\theta) < 1$ for $0 < \theta < 1$, and that

(ii) $f(\theta) > 1$ for $\theta < 0$. Therefore $\theta = 0$ is neither a maximizer or minimizer of f. It is a horizontal point of inflection of $f(\theta)$.

Functions of One Variable

Example

Note that

$$f'(\epsilon) = 12\epsilon^2(\epsilon - 1) < 0, \quad (1)$$

$$f'(-\epsilon) = 12(-\epsilon)^2(-\epsilon - 1) < 0, \quad (2)$$

so $\theta = 0$ is a critical point but not a turning point.

Remark: A turning point is a point at which the derivative changes sign. A turning point may be either a local minimum or a local maximum. If the function is differentiable, then a turning point is a stationary point; however not all stationary points are turning points.

Check this out - Anyone training ML models should read this!

Functions of One Variable

Our next objective is to extend the results to functions of more than one variable by combining calculus and linear algebra.

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%

Homework 1 - Check Blackboard - Due 28-**Sep**

Symmetric and Positive Definite Matrices

Symmetric matrices have several special properties, particularly with respect to their eigenvalues and eigenvectors. They also play an important role in optimization. For example, the Hessian matrix $H(\theta) = \nabla^2 f(\theta)$ is symmetric.

Symmetric and Positive Definite Matrices

Theorem (Proposition: Spectral decomposition)

Let S^n be the space of $n \times n$ (real) symmetric matrices and let $A \in S^n$. Then

1. $\lambda(A)$ *are real*
2. *A can be decomposed in the form $A = PDP^T$ where P is an orthogonal matrix, i.e., $P^TP = PP^T = I$, and $D = diag(\lambda_1, \ldots, \lambda_n)$.*
3. *Suppose that the eigenvectors v_i are normalized, i.e. $\|v_i\| = 1 \; \forall i = 1, \ldots, n$. Then $A = \sum_{i=1}^{n} \lambda_i v_i v_i^T$, where λ_i is the eigenvalue corresponding to v_i.*

Symmetric and Positive Definite Matrices

Definition (Positive definiteness)

A matrix $A \in S^n$ is said to be positive definite if $x^T Ax > 0 \ \forall x \in \mathbb{R}^n$ and $x \neq 0$. This is denoted $A \succ 0$. If $x^T Ax \geq 0 \ \forall x \in \mathbb{R}^n$ then A is said to be nonnegative definite or positive semidefinite. This is denoted $A \succeq 0$.

Proposition

For $A \in S^n$,

$$A \succ 0 \iff \lambda_i > 0, \ \forall \lambda_i \in \lambda(A)$$
$$A \succeq 0 \iff \lambda_i \geq 0, \ \forall \lambda_i \in \lambda(A).$$

Symmetric and Positive Definite Matrices

Example

For

$$A = \begin{bmatrix} 3 & -2 & 2 \\ -2 & 7 & -2 \\ 2 & -2 & 3 \end{bmatrix}, \text{ we have } \lambda(A) = \{1, 3, 9\},$$

so that $A \succ 0$. We see that $\Delta_1 = 3$, $\Delta_2 = \det \begin{bmatrix} 3 & -2 \\ -2 & 7 \end{bmatrix} = 17$, and $\Delta_3 = 27$, so all the principal minors are positive.

Symmetric and Positive Definite Matrices

Remark:

1. $\Delta_k \geq 0 \; \forall k = 1, \ldots, n \not\Rightarrow A \succeq 0$
2. $\Delta_1 > 0$, $\Delta_2 > 0$, **. . . ,** $\Delta_{n-1} > 0$, $\Delta_n = 0 \implies A \succeq 0$.
3. **If** $(-1)^k \Delta_k > 0$ **for** $k = 1, \ldots, n-1$ **while** $\Delta_n = 0$ **then** $A \preceq 0$.

Symmetric and Positive Definite Matrices

Example

For

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & \frac{1}{2} \end{bmatrix}$$

we get $\Delta_1 \geq 0$, $\Delta_2 \geq 0$, $\Delta_3 \geq 0$, but A is not positive semidefinite. For $x = (1, 1, -2)^T$, we get $x^T A x = -2 < 0$. Note that $\lambda(A) = -0.3508, 0, 2.8508$, so A is indefinite.