Foundations of Machine Learning

Brett Bernstein

August 22, 2018

Lecture 4: Concept Checks

Convexity

Optional Learning Objectives

Convex optimization and Lagrangian duality will not covered on the midterm exam, so in some sense these objectives are optional.

- Define a convex set, a convex function, and a strictly convex function. (Don't forget that the domain of a convex function must be a convex set!)
- For an optimization problem, define the terms feasible set, feasible point, active constraint, optimal value, and optimal point.
- Give the form for a general inequality-constrained optimization problem (there are many ways to do this, but our convention is to have inequality constraints of the form $f_i(x) \le 0$).
- Define the Lagrangian for this optimization problem, and explain how the Lagrangian encodes all the information in the original optimization problem.
- Write the primal and dual optimization problem in terms of the Lagrangian.

Convexity Concept Check Problems

1. If $A, B \subseteq \mathbb{R}^n$ are convex, then $A \cap B$ is convex.

   Solution. Let $x, y \in A \cap B$ and $t \in (0,1)$. Since A, B are convex, we have
   $$(1-t)x + ty \in A \quad \text{and} \quad (1-t)x + ty \in B.$$
   Thus $(1-t)x + ty \in A \cap B$.

2. Let $f, g : \mathbb{R}^n \to \mathbb{R}$ be convex. Show that $af + bg$ is convex if $a, b \ge 0$.

Solution. Let $x, y \in \mathbb{R}^n$ and $\theta \in (0,1)$. Then

$$\begin{aligned}(af+bg)((1-\theta)x+\theta y) &= af((1-\theta)x+\theta y)+bg((1-\theta)x+\theta y)\\ &\le a[(1-\theta)f(x)+\theta f(y)]+b[(1-\theta)g(x)+\theta g(y)]\\ &= (1-\theta)(af+bg)(x)+\theta(af+bg)(y).\end{aligned}$$

3. Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and differentiable. Prove that if $\nabla f(x) = 0$ then x is a global minimizer.

   Solution. Suppose $\nabla f(x) = 0$. The gradient (or first-order) characterization of convexity says
$$f(y) \ge f(x) + \nabla f(x)^T(y-x)$$
   for all y. If $\nabla f(x) = 0$ then this says $f(y) \ge f(x)$ for all x.

4. Prove that if $f : \mathbb{R}^n \to \mathbb{R}$ is strictly convex and x is a global minimizer, then it is the unique global minimizer.

   Solution. Suppose y is also a global minimizer with $y \ne x$. Then
$$f((y+x)/2) < f(y)/2 + f(x)/2 = f(x)$$
   contradicting the fact that $f(x)$ was a global minimizer.

5. Prove that any affine function $f : \mathbb{R}^n \to \mathbb{R}$ is both convex and concave.

   Solution. Recall that f has the form $f(x) = w^T x + b$ where $w \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Then, for $x, y \in \mathbb{R}^n$ and $\theta \in (0,1)$,
$$f((1-\theta)x+\theta y) = w^T((1-\theta)x+\theta y)+b = (1-\theta)(w^T x+b)+\theta(w^T y+b) = (1-\theta)f(x)+\theta f(y).$$
   This shows f is convex. But the same holds if we replace w with $-w$ and b with $-b$. Hence f is also concave.

6. Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and let $g : \mathbb{R}^m \to \mathbb{R}^n$ be affine. Then $f \circ g$ is convex.

   Solution. Write $g(x) = Ax + b$ where $A \in \mathbb{R}^{n\times m}$ and $b \in \mathbb{R}^n$. For $x, y \in \mathbb{R}^m$ and $t \in (0,1)$ we have
$$\begin{aligned}f(g((1-t)x+ty)) &= f((1-t)(Ax+b)+t(Ay+b))\\ &\le (1-t)f(Ax+b)+tf(Ay+b)\\ &= (1-t)f(g(x))+tf(g(y)).\end{aligned}$$

7. ($\star\star$)

   (a) Let $f : \mathbb{R} \to \mathbb{R}$ be convex. Show that f has one-sided left and right derivatives at every point.

(b) Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. Show that f has one-sided directional derivatives at every point.

(c) Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. Show that if x is not a minimizer of f then f has a descent direction at x (i.e., a direction whose corresponding one-sided directional derivative is negative).

Solution. We first prove the following lemma.

Lemma 1. *If $f : \mathbb{R} \to \mathbb{R}$ is convex and $x < y < z$ then*

$$\frac{f(y) - f(x)}{y - x} \le \frac{f(z) - f(x)}{z - x}.$$

Proof. Let $t \in (0, 1)$ satisfy $(1 - t)x + tz = y$. By convexity we have

$$f(y) = f((1 - t)x + tz) \le (1 - t)f(x) + tf(z)$$

giving

$$\frac{f(y) - f(x)}{y - x} \le \frac{(1 - t)f(x) + tf(z) - f(x)}{(1 - t)x + tz - x} = \frac{t(f(z) - f(x))}{t(z - x)} = \frac{f(z) - f(x)}{z - x}.$$

□

(a) For the right derivative, we will show

$$\lim_{y \downarrow x} \frac{f(y) - f(x)}{y - x} = \inf_{y > x} \frac{f(y) - f(x)}{y - x} =: L.$$

Fix $\epsilon > 0$ and choose $y' > x$ so that

$$\frac{f(y') - f(x)}{y' - x} < L + \epsilon.$$

Letting $\delta = y' - x$, the lemma shows that

$$\frac{f(y) - f(x)}{y - x} < L + \epsilon$$

for any $y < x + \delta$ proving the limit exists.

For the left derivative, we could repeat the above, or note that $g(t) = 2x - t$ is affine, so $f \circ g$ is convex. By the above

$$\lim_{y \downarrow x} \frac{f(g(y)) - f(g(x))}{y - x} = \lim_{y \downarrow x} \frac{f(2x - y) - f(x)}{y - x} = \lim_{h \downarrow 0} \frac{f(x - h) - f(x)}{h}$$

exists, where $h = y - x$. This proves the left derivative exists as well.

(b) Fix $x, v \in \mathbb{R}^n$ and let $g : \mathbb{R} \to \mathbb{R}^n$ be defined by $g(t) = x + tv$. Then $f \circ g$ is convex, and thus the previous part applies. But the right derivative of g at 0 is the one-sided directional derivative of f at x in the direction v:

$$\lim_{h \downarrow 0} \frac{f(g(h)) - f(g(0))}{h} = \lim_{h \downarrow 0} \frac{f(x + hv) - f(x)}{h}.$$

(c) Let y be a minimizer of f and let $g(t) = x + t(y - x)$. By the arguments in the first part above, the value

$$\frac{f(g(1)) - f(g(0))}{1 - 0} = f(y) - f(x) < 0$$

is an upper bound on the right derivative of g at 0. But this is a directional derivative, by the argument in the second part above.

Convex Optimization Problems

1. Suppose there are mn people forming m rows with n columns. Let a denote the height of the tallest person taken from the shortest people in each column. Let b denote the height of the shortest person taken from the tallest people in each row. What is the relationship between a and b?

   Solution. Let H_{ij} denote the height of the person in row i and column j. Then

$$a = \max_j \min_i H_{ij} \leq \min_i \max_j H_{ij} = b,$$

   by the max-min inequality.

2. Let $x_1, \ldots, x_n \in \mathbb{R}^d$ be given data. You want to find the center and radius of the smallest sphere that encloses all of the points. Express this problem as a convex optimization problem.

   Solution.

$$\begin{array}{ll} \text{minimize}_{r,c} & r \\ \text{subject to} & \|x_i - c\|_2 \leq r \quad \text{for } i = 1, \ldots, n. \end{array}$$

   This problem is convex since norms are convex, so $f_i(c) = \|x_i - c\|_2$ is convex (composition of convex with affine).

3. Suppose $x_1, \ldots, x_n \in \mathbb{R}^d$ and $y_1, \ldots, y_n \in \{-1, 1\}$. Here we look at y_i as the label of x_i. We say the data points are linearly separable if there is a vector $v \in \mathbb{R}^d$ and $a \in \mathbb{R}$ such that $v^T x_i > a$ when $y_i = 1$ and $v^T x_i < a$ for $y_i = -1$. Give a method for determining if the given data points are linearly separable.

Solution. Solve the hard-margin SVM problem

$$\begin{array}{ll} \text{minimize}_{w,b} & \|w\|_2^2 \\ \text{subject to} & y_i(w^T x_i + b) \geq 1 \quad \text{for all } i = 1, \ldots, n. \end{array}$$

If the resulting problem is feasible, then the data is linearly separable.

4. Consider the Ivanov form of ridge regression:

$$\begin{array}{ll} \text{minimize} & \|Ax - y\|_2^2 \\ \text{subject to} & \|x\|_2^2 \leq r^2, \end{array}$$

where $r > 0$, $y \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times n}$ are fixed.

(a) What is the Lagrangian?

(b) What do you get when you take the supremum of the Lagrangian over the feasible values for the dual variables?

Solution.

(a) $L(x, \lambda) = \|Ax - y\|_2^2 + \lambda(\|x\|_2^2 - r^2)$. Note that this is a shifted version of the Tikhonov objective.

(b)
$$\sup_{\lambda \succeq 0} L(x, \lambda) = \begin{cases} +\infty & \text{if } \|x\|_2^2 > r^2, \\ \|Ax - y\|_2^2 & \text{otherwise.} \end{cases}$$

Note that the original Ivanov minimization is then just

$$\inf_x \sup_{\lambda \succeq 0} L(x, \lambda).$$