

# NYU Center for Data Science: DS-GA 1003

## Machine Learning and Computational Statistics (Spring 2018)

Brett Bernstein

February 12, 2018

**Instructions:** Following most lab and lecture sections, we will be providing concept checks for review. Each concept check will:

- List the lab/lecture learning objectives. You will be responsible for mastering these objectives, and demonstrating mastery through homework assignments, exams (midterm and final), and on the final course project.
- Include concept check questions. These questions are intended to reinforce the lab/lectures, and help you master the learning objectives.

You are strongly encourage to complete all concept check questions, and to discuss these (and related) problems on Piazza and at office hours. However, problems marked with a (★) are considered optional.

## Lecture 4: Concept Checks

### Convexity

#### Learning Objectives

- 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  $f_i(x) \leq 0$ ).
- 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.
2. Let  $f, g : \mathbb{R}^n \rightarrow \mathbb{R}$  be convex. Show that  $af + bg$  is convex if  $a, b \geq 0$ .
3. Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be convex and differentiable. Prove that if  $\nabla f(x) = 0$  then  $x$  is a global minimizer.
4. Prove that if  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is strictly convex and  $x$  is a global minimizer, then it is the unique global minimizer.
5. Prove that any affine function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is both convex and concave.
6. Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be convex and let  $g : \mathbb{R}^m \rightarrow \mathbb{R}^n$  be affine. Then  $f \circ g$  is convex.
7. (★★)
  - (a) Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be convex. Show that  $f$  has one-sided left and right derivatives at every point.
  - (b) Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be convex. Show that  $f$  has one-sided directional derivatives at every point.
  - (c) Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  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).

## 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  $b$ ?
2. Let  $x_1, \dots, x_n \in \mathbb{R}^d$  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.
3. Suppose  $x_1, \dots, x_n \in \mathbb{R}^d$  and  $y_1, \dots, y_n \in \{-1, 1\}$ . 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.
4. Consider the Ivanov form of ridge regression:

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

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?