Lecture 13

1. Applications of Linear Programming, Part II

1.1. **Optimal classifiers.** Suppose we are given two finite sets $P, Q \subset \mathbb{R}^s$. These are our 'training' data, from which we want to find an optimal classifier $f\colon \mathbb{R}^s \to \mathbb{R}$ that distinguishes P from Q, with the goal of later using it to decide if new points $x \in \mathbb{R}^s$ should be grouped with P or Q.

A classifier that separates P and Q is a function $f\colon \mathbb{R}^s \to \mathbb{R}$ such that

$$f(x) > 0 \text{ for all } x \in P \qquad \text{and} \qquad f(x) < 0 \text{ for all } x \in Q. \tag{1}$$

Given $\delta \geq 0$, we say that such a classifier is δ-good if it satisfies

$$f(x) \geq \delta \text{ for all } x \in P \qquad \text{and} \qquad f(x) \leq -\delta \text{ for all } x \in Q. \tag{2}$$

The larger $\delta > 0$ is, the better a δ-good classifier is, since it affords us more room separating the sets P and Q. Geometrically, if $f\colon \mathbb{R}^s \to \mathbb{R}$ is a δ-good classifier, then not only the hypersurface $\{x \in \mathbb{R}^s : f(x) = 0\}$ separates P and Q, but we can fit an open band of width 2δ around that hypersurface without touching either P or Q. Thus, we consider a classifier $f\colon \mathbb{R}^s \to \mathbb{R}$ *optimal* if it is δ-good for the largest possible $\delta > 0$.

For now, let us consider *affine* classifiers, i.e., affine maps $f\colon \mathbb{R}^s \to \mathbb{R}$, $f(x) = a^T x + b$.

Exercise 1. Give an example of sets $P, Q \subset \mathbb{R}^2$ that cannot be separated by an affine classifier.

Solution to Exercise 1. $P = \{(0,0), (1,1)\}$, $Q = \{(1,0), (0,1)\}$.

Exercise 2. Give an example of finite sets $P, Q \subset \mathbb{R}^2$ with $|P| = |Q| = 2$ that can be separated by an affine classifier, and find the largest $\delta > 0$ such that δ-good affine classifiers exist.

Solution to Exercise 2. Consider the sets $P = \{(0,1), (1,1)\}$ and $Q = \{(0,0), (1,0)\}$. Geometrically, it seems clear that the largest possible δ is $\delta = 1/2$, with optimal affine classifier $f(x) = 1/2$. In order to prove this, we set up an LP as follows:

$$\begin{aligned} \max \quad \delta \quad \text{s.t.} \quad a + b - 1 &\leq -\delta, \\ b - 1 &\leq -\delta, \\ a + b &\geq \delta, \\ b &\geq \delta. \end{aligned}$$

From the last constraint, it actually follows that $b \geq 0$. However, a is unconstrained. Thus, we replace a with $a = a_1 - a_2$ where $a_1 \geq 0$ and $a_2 \geq 0$. Then, adding slack variables, subtracting surplus variables and adding artificial variables, we end up with the following LP in equational form:

$$\begin{aligned} \max \quad \delta \quad \text{s.t.} \qquad a_1 - a_2 + b + \delta + s_1 &= 1, \\ b + \delta + s_2 &= 1, \\ a_1 - a_2 + b - \delta - s_3 + q_1 &= 0, \\ b - \delta - s_4 + q_2 &= 0. \end{aligned}$$

The optimal value for the above LP is $\delta = \frac{1}{2}$, achieved at $(a, b) = (0, \frac{1}{2})$, as shown using the two-phase simplex method in `lecture13.nb`.

Exercise 3. Suppose NYC wants to build a computer-controlled trap to be placed in public parks to captures rats, but not squirrels.[1] In order to separate rats from squirrels, a scale and an imaging device in the trap measure the animals' weight and shadow area, according to the recommendation of biologists. We are given the following "training data" to distinguish rats from squirrels:

[1]This example is modified from Sec. 2.5 in "Understanding and Using Linear Programming", by Jiri Matousek and Bernd Gärtner (Springer).

	weight	area
Squirrels	1	1.5
	2	2.2
	2.3	2
Rats	1.4	1
	2.2	1.5
	3	1.7

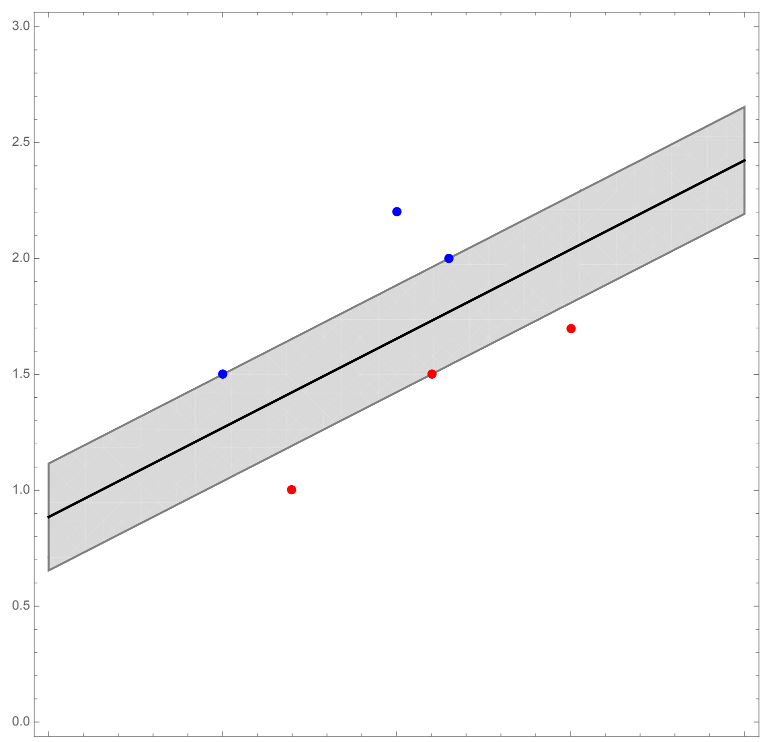
Find an optimal affine classifier $f\colon \mathbb{R}^2 \to \mathbb{R}$ to be programmed into this computer-controlled trap, so that an animal with $x = (\text{weight,area})$ is captured if $f(x) > 0$, presumed to be a rat, and released if $f(x) < 0$, presumed to be a squirrel.

Solution to Exercise 3. To simplify notation (and make it geometrically more intuitive), instead of a general $f(x) = a_1x_1 + a_2x_2 + b$, we fix $a_2 = -1$ and consider affine classifiers $f(x) = a\,x_1 - x_2 + b$, so that $f(x) = 0$ is the line $x_2 = a\,x_1 + b$ in the (x_1, x_2)-plane. Optimal classifiers must solve

$$\begin{aligned} \max \quad \delta \quad \text{s.t.} \quad a + b - 1.5 &\le -\delta, \\ 2a + b - 2.2 &\le -\delta, \\ 2.3a + b - 2 &\le -\delta, \\ 1.4a + b - 1 &\ge \delta, \\ 2.2a + b - 1.5 &\ge \delta, \\ 3a + b - 1.7 &\ge \delta. \end{aligned}$$

The optimal solution is $a = 0.384615$, $b = 0.884615$, which attains the largest possible $\delta = 0.230769$, so we use the affine classifier $f(x) = 0.384615x + 0.884615$, see file `lecture13.nb` for details.

In the graph below, data points for squirrels and rats are blue and red, respectively, the black line is given by $f(x) = 0$ and the gray band has (vertical) width $\delta = 0.230769$.



Several generalizations of the above are possible. First, our data points may be contained in a more complicated set X, which is not necessarily an Euclidean space $\mathbb{R}^s$. Second, since the evaluation map $x \mapsto f(x)$ is obviously linear in $f \colon X \to \mathbb{R}$, we can use a linear program to find optimal classifiers among linear combinations of a (finite) set of functions $\{f_1, \dots, f_n\}$, which need not be affine. Once again, the *optimal* classifier is $f = a_1 f_1 + \cdots + a_n f_n$ which maximizes $\delta > 0$ such that $f(x) \geq \delta$ for all $x \in P$ and $f(x) \leq -\delta$ for all $x \in Q$. The particular case of affine classifiers, discussed above, corresponds to $X = \mathbb{R}^s$ and $f_j(x) = e_j^T x$, for $j = 1, \dots, s$, and $f_{s+1}(x) = 1$.

1.2. **Largest disk in a convex polygon.** Read Sec. 2.6 in "Understanding and Using Linear Programming", by Jiri Matousek and Bernd Gärtner (Springer) and solve:

Exercise 4. Find the largest disk inscribed in the polygon in $\mathbb{R}^2$ determined by the inequalities $x_1 + x_2 \leq 3$, $x_1 - x_2 \leq 1$, $x_1 \geq 0$, $x_2 \geq 0$.

Solution to Exercise 4. The center of the disk is $\left(\frac{2\sqrt{2}}{2+\sqrt{2}}, 1\right)$ and radius is $\frac{2\sqrt{2}}{2+\sqrt{2}}$.