1. Here is a linear program, over variables $x \in \mathbb{R}^n$ and $v \in \mathbb{R}$:

$$\min v$$

$$-b_i + \sum_{j=1}^{n} a_{ij} x_j \leq v, \quad i = 1, 2, \ldots, m$$

$$b_i - \sum_{j=1}^{n} a_{ij} x_j \leq v, \quad i = 1, 2, \ldots, m$$

2. (a) Let $K$ denote the intersection of halfspaces given by $w_1, w_2, \ldots \in \mathbb{R}^d$ and $b_1, b_2, \ldots \in \mathbb{R}$:

$$K = \bigcap_i \{x : w_i \cdot x \leq b_i\}.$$  

For any $x, y \in K$ and $0 < \theta < 1$,

$$w_i \cdot (\theta x + (1 - \theta)y) = \theta w_i \cdot x + (1 - \theta) w_i \cdot y \leq \theta b_i + (1 - \theta) b_i = b_i, \quad \text{for } i = 1, 2, \ldots.$$ 

Therefore, $\theta x + (1 - \theta)y \in K$; and $K$ is a convex set.

(b) The unit ball in $\mathbb{R}^d$ can be written as

$$\bigcap_{\|w\|=1} \{x : w \cdot x \leq 1\}.$$ 

3. $P_1$ and $P_2$ are polyhedra that are intersections of finitely many halfspaces. Let the halfspaces for $P_1$ be given by $u_1, \ldots, u_m \in \mathbb{R}^d$ and $b_1, \ldots, b_m \in \mathbb{R}$:

$$P_1 = \bigcap_{i=1}^{m} \{x : u_i \cdot x \leq b_i\}.$$ 

Likewise, let $P_2$ be given by $v_1, \ldots, v_n \in \mathbb{R}^d$ and $c_1, \ldots, c_n \in \mathbb{R}$:

$$P_2 = \bigcap_{i=1}^{n} \{x : v_i \cdot x \leq c_i\}.$$ 

We wish to find the point $x_1 \in P_1$ and $x_2 \in P_2$ that are closest to one another. Let us write $z = x_1 - x_2$. Here is the optimization problem:

$$\min \|z\|^2$$

$$u_i \cdot x_1 \leq b_i, \quad i = 1, 2, \ldots, m$$

$$v_i \cdot x_2 \leq c_i, \quad i = 1, 2, \ldots, n$$

$$z = x_1 - x_2$$

The constraints are all linear, and the objective function is convex, so this is a convex optimization problem.

(a) There are as many disjunctions as there are subsets of features, so $|\mathcal{H}| = 2^d$.

(b) The true error of $h$ can be bounded thus, with probability at least $1 - \delta$:

$$\text{err}(h) \leq \frac{1}{n} \ln \frac{|\mathcal{H}|}{\delta} = \frac{1}{n} \left( d \ln 2 + \ln \frac{1}{\delta} \right).$$

(c) $|\mathcal{H}_k| \leq d^k$, so we get

$$\text{err}(h) \leq \frac{1}{n} \ln \frac{|\mathcal{H}|}{\delta} = \frac{1}{n} \left( k \ln d + \ln \frac{1}{\delta} \right).$$

5. By the central limit theorem, $\hat{p}$ follows roughly a $N(3/4, 1/1600)$ distribution. With 95% probability, $\hat{p}$ will fall within 2 standard deviations of its mean, that is, in the interval $[0.7, 0.8]$.

6. VC dimension.

(a) The class $\mathcal{H}$ of intervals on the real line shatters any set of two distinct points: it can realize all four labelings of these points. But it cannot shatter any set of three points, because it cannot label the middle one 0 while making the other two 1. Therefore $\text{VC}(\mathcal{H}) = 2$.

(b) The class $\mathcal{H}$ of axis-aligned rectangles in the plane shatters the set $\{(0, 1), (0, -1), (1, 0), (-1, 0)\}$: all 16 labelings can be realized. But it cannot shatter any set of five points. To see this, pick any $x_1, \ldots, x_5 \in \mathbb{R}^2$. One of them must lie in the bounding box of the other four points; say $x_5$ lies in the bounding box of $x_1, x_2, x_3, x_4$. Then we cannot realize the labeling $y_1 = y_2 = y_3 = y_4 = 1$ and $y_5 = 0$. Thus $\text{VC}(\mathcal{H}) = 4$.

7. Isotonic regression.

(a) Here’s a monotonic function that goes through four of the points.

![Figure 1: Sketch of $f(x)$](image)
(b) We can write the least-squares isotonic regression problem as follows:

$$\min L(f) = \sum_{i=1}^{n} (y_i - f_i)^2$$

$$f_i - f_{i+1} \leq 0 \quad \text{for } i = 1, 2, \ldots, n - 1$$

The constraints are linear in $f$, and the objective function $L(f)$ is convex: its Hessian is $H(f) = 2I$, which is positive semidefinite. Therefore, the problem above is a convex problem.

(c) When the pool-adjacent-violators algorithm is applied to the given set of six points, the final adjusted values are:

$$(1, 22/3), (2, 22/3), (3, 22/3), (4, 41/3), (5, 41/3), (6, 41/3).$$