

**DATA 1010**  
**IN-CLASS EXERCISES**  
**SAMUEL S. WATSON**  
**12 NOVEMBER 2018**

**Problem 1**

Find the distance from the plane  $3x + 2y + z = 6$  to the point  $P = (4, 7, 1)$ .

**Problem 2**

Find the distance from the hyperplane  $\{\mathbf{x} \in \mathbb{R}^n : \boldsymbol{\beta} \cdot \mathbf{x} - \alpha = 0\}$  to the point  $\mathbf{x}$ .

**Problem 3**

Simulate data for a binary classification problem in the plane for which the two classes can be separated by a line. Write a Julia function for finding the thickest slab which separates the two classes.

**Problem 4**

Now suppose that the data are not separable by a plane. Explain why

$$L(\boldsymbol{\beta}, \alpha) = \lambda |\boldsymbol{\beta}|^2 + \frac{1}{n} \sum_{i=1}^n [1 - y_i(\boldsymbol{\beta} \cdot \mathbf{x}_i - \alpha)]_+$$

is a reasonable quantity to minimize. (Note:  $u_+$  means  $\max(0, u)$ , and  $\lambda$  is a parameter of the loss function).

**Problem 5**

Simulate some overlapping data and minimize the loss function given in Problem 4. Choose the value of  $\lambda$  using cross-validation.