How Does the Loss Determine the Center?

Due: Wednesday, October 7 at 11:59pm

Instructions: Please submit your *handwritten* solutions as a single PDF on Gradescope.

An array of sensors reports $n \geq 1$ readings $y_1, \ldots, y_n \in \mathbb{R}$. We want one number $c \in \mathbb{R}$ to represent them. Different ways of measuring error can favor different notions of a center.

Define the residual vector $\mathbf{r}(c) = (y_1 - c, \ldots, y_n - c)^\top \in \mathbb{R}^n$. Consider these four costs:

$$\|\mathbf{r}(c)\|_0 = \#\{i : y_i \neq c\}, \qquad \|\mathbf{r}(c)\|_1 = \sum_{i=1}^{n} |y_i - c|,$$

$$\|\mathbf{r}(c)\|_2 = \left(\sum_{i=1}^{n} (y_i - c)^2\right)^{1/2}, \qquad \|\mathbf{r}(c)\|_\infty = \max_{1 \leq i \leq n} |y_i - c|.$$

Here $\#$ counts elements. The conventional notation $\|\cdot\|_0$ counts nonzero coordinates. It is not a norm. This mismatch count is minimized by any most frequent reading, called a mode, because it equals n minus the number of readings equal to c.

1. Absolute error adds the distances to the readings. Determine the complete set of constants c minimizing $\|\mathbf{r}(c)\|_1$.
2. Euclidean error combines the squared distances. Determine the complete set of constants c minimizing $\|\mathbf{r}(c)\|_2$.
3. Maximum error measures the distance to the farthest reading. Determine the complete set of constants c minimizing $\|\mathbf{r}(c)\|_\infty$.

Takeaway: *The same readings can have different best centers: a mode, a median, the mean, or the midpoint of the extremes. The answer depends on how we measure error. Choosing a loss decides what we mean by a good prediction.*