1 Regression

1.1 Introduction

In this section we present a remedial solution to combine our regression models with the pointwise-estimation baseline model, and a more principled solution. Since we are combining the models, our gold standards will now be the original annotated sets composed of with size between two to ten.

1.2 Remedial Solution

First we give brief reminder for our baseline model. We defined two possible events: $\boldsymbol{\Omega} = \{s < t, s > t\}$, and after observing a sequence of comparisons between s and t: $\boldsymbol{S} = \{s < t, s < t, \ldots, s > t \ldots\}$, we can ask what is the probability that the next element we will observe is $s < t$. This is a Bernoulli distribution with parameter p and it is well known that the most likely p is simply:

$$\boldsymbol{Pr}[s < t] = \frac{|\{s < t \in \boldsymbol{S}\}|}{|\boldsymbol{S}|}.$$

In the baseline, if $\boldsymbol{S}$ is empty then we defaulted to $\boldsymbol{Pr}[s < t] = \frac{1}{2}$.

Now we present the remedial solution. Recall in the previous chapter we defined this probability value for the $\hat{y}$ output by elastic net regression:

$$\boldsymbol{Pr}[s < x] = \begin{cases} \frac{1}{2} + \epsilon & \hat{y} < \delta \\ \frac{1}{2} - \epsilon & otherwise, \end{cases}$$

while we used the actual probabilty value p output by the logistic regression model. In the remedial solution, we use the elastic definition defined above, and in the case of logistic regression, we actually discard the value of p and define:

$$\boldsymbol{Pr}[s < x] = \begin{cases} \frac{1}{2} + \epsilon & p > \frac{1}{2} \\ \frac{1}{2} - \epsilon & otherwise. \end{cases}$$

This captures our intuition that the prediction output by the model is less accruate than that of the actual data. Results are displayed below.