

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: $\Omega = \{s < t, s > t\}$, and after observing a sequence of comparisons between s and t : $\mathcal{S} = \{s < t, s < t, \dots, s > t \dots\}$, 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:

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

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

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:

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

while we used the actual probability value p output by the logistic regression model. In the remedial solution, we use the elastic definition defined above, and in

	Elastic Net Regression		l_1 -Logistic Regression	
Gold Set	Pairwise	Avg. τ	Pairwise	Avg. τ
BCS	90.0%	0.81	93.0%	0.85
Turk	75.0%	0.62	74.0%	0.61
Turk no-tie	81.0%	0.63	81.0%	0.62
Mohit	74.0%	0.61	74.0%	0.61
Mohit no-tie	76.0%	0.52	76.0%	0.53

Table 1: . Results for the two best models combined with pointwise estimation baseline in the remedial fashion. Note how two models performs comparable across all gold sets. In addition, not the gold clusters with no ties enjoyed a higher pairwise accuracy but suffer a lower τ value.

the case of logistic regression, we actually discard the value of p and define:

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

This captures our intuition that the prediction output by the model is less accurate than that of the actual data. Additionally, we also constructed a version of the Turk and Mohit’s clusters where ties are removed. We reasoned that since our models are designed to predict ordering, while ties can be interpreted as synonyms, clusters generated without ties may give a more “fair” representation of how well the models perform. Results are displayed below.

1.3 Solution with Beta Prior

In this section we provide a more formal variant of the remedial solution. The heart of the of the problem is that we have some prior belief about the likelihood that one adjective is weaker than another, and an updated belief after observing some data, be it direct comparison or estimation from a model. Since we are modeling each

edge a Bernoulli variable with parameter p , the prior is then a distribution over the Bernoulli p , this is the Beta distribution. In this next few paragraphs, we give a brief overview of Beta-Binomial model, in particular how it applies to our problem.