

Zero probabilities for test data

Toy train corpus:

This is the house that Jack built.

Toy test corpus:

This is *Jack*.

What's the perplexity of the Bigram LM?

$$p(Jack \mid is) = \frac{c(is \ Jack)}{c(is)} = 0$$

$$p(\mathbf{w}_{\text{test}}) = 0$$

$$\mathcal{P} = \inf$$

