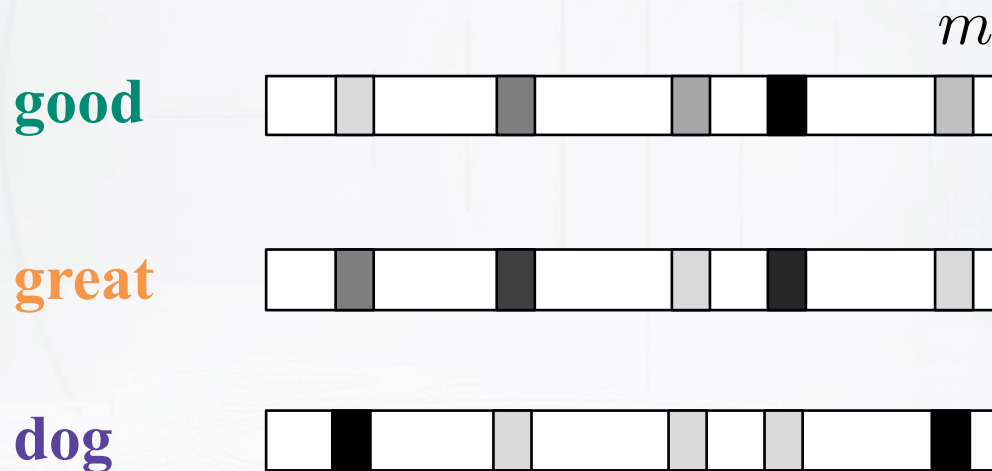


How to generalize better

- Learn **distributed representations** for words
- Express probabilities of sequences in terms of these distributed representations and learn parameters



$C^{|V| \times m}$ – matrix of distributed word representations.