

log loss encourages unbiased prediction

- ▶ Suppose the source is random and the probability of the next outcome is $p(c_t | c_1, c_2, \dots, c_{t-1})$
- ▶ Then the prediction that minimizes the log loss is $p(c_t | c_1, c_2, \dots, c_{t-1})$.