Distances and Similarities

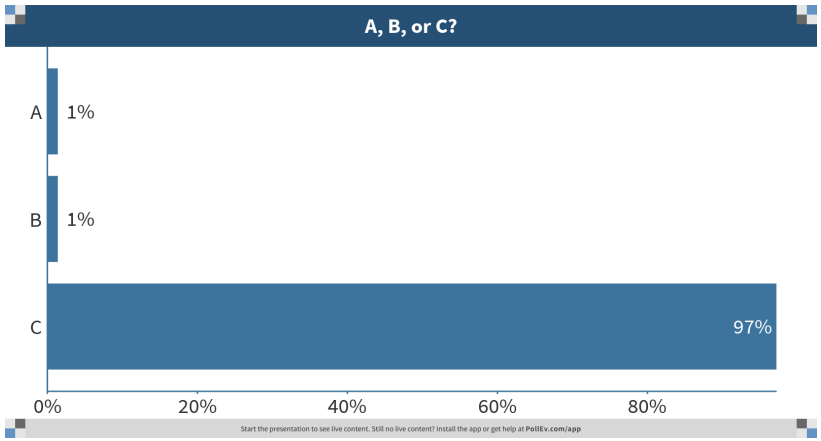
- **Distances**
  - What distances have we used?
  - What properties do they have?
- **Similarities**
  - What similarity measures have we used?
  - What properties do they have

KL-Divergence

A) Distance

B) Similarity

C) Neither



$$D_{\mathrm{KL}}(P\|Q) = -\sum_{i} P(i)\,\log\frac{Q(i)}{P(i)},$$

Kullback Leibler divergence

- ***P*** = 'true' distribution
- ***Q*** = alternative distribution that is used to encode data
- **KL** divergence is the expected extra # of bits per point that must be transmitted using ***Q*** *instead of* ***P*** *if the data comes from* ***P***

$$D_{KL}(P \,||\, Q) = \sum_i P(x_i) \log (P(x_i)/Q(x_i))$$

$$= -\sum_i P(x_i) \log Q(x_i) + \sum_i P(x_i) \log P(x_i)$$

$$= H(P,Q) - H(P)$$

= Cross-entropy - entropy

- **Measures how different the two distributions are**

KL divergence as error

- Given a label y=a for a categorical variable which is one of k outcomes, P is a unit vector (a "one hot" encoding).
- The output of a logistic regression or neural net, or any softmax function is a distribution Q over the k possible outcomes
- The KL divergence is a good loss function

$$D_{KL}(P \,||\, Q) = \sum_k P(y=k) \log (P(y=k)/Q(y=k)) = -\log(Q(y=a))$$

What is the loss if Q(y=a)=1?

if Q(y=a)=0?

KL divergence as info gain

- **The KL divergence of the posterior measures the information gain expected from query (x'):**

$$D(p(\theta \mid x, x') \,||\, p(\theta \mid x))$$

- **Goal: choose a query that *maximizes* the KL divergence between the updated posterior probability and the current posterior probability**
  - This represents the largest expected information gain

KL divergence properties

- **Measures how well a probability distribution Q approximates a distribution P (the "truth")**
- **Divergence *0* if and only if *P* and *Q* are equal*:***
  - $D(P||Q) = 0 \text{ iff } P = Q$
- **Non-symmetric*: $D(P||Q) \neq D(Q||P)$***
- **Non-negative*: $D(P||Q) \geq 0$***
- **Does not satisfy triangle inequality**
  - $D(P||Q) \leq D(P||R) + D(R||Q)$

Not a distance metric

Questions?

Top