information theory

colin shaw
01.21.2016

1. **introduction**
    a. welcome to computer science club
    b. thank you to revunit for sponsoring
    c. what is it about
        i. practice presenting
        ii. demonstrating competencies
        iii. promoting the community
    d. nod to addam hardy and his id3 presentation
    e. introduce tyler
    f. general overview of presentation

2. **a bit of history - a quick overview**
    a. ralph hartley (1928) - same year as first television station
        i. electrical engineering
        ii. resolvable symbols
        iii. rate of transmission
        iv. basic idea of what we now call entropy
        v. basic idea of channel capacity and noise
    b. claude shannon (1948)
        i. bell labs
        ii. greatly expanded on hartley's ideas
            1. more formal definition of entropy
            2. mutual information (will see later)
            3. channel capacity
            4. noisy channel
    c. what fields this has affected - just some examples
        i. communication encoding
        ii. validation and integrity (crc, hashes)
        iii. cryptography (rng fitness)
        iv. machine learning (decision trees)
        v. linguistics
        vi. neurobiology

3. **preliminary concepts**
    a. motivating data examples
        i. analog communication signals
            1. signal to noise
            2. studies of intelligibility
        ii. alphabet
            1. naive 26 character symbols
            2. higher order english language formatting (words, sentences, etc.)
        iii. note: descriptive not prescriptive
        iv. note: stationary - does not depend on preceding information
    b. entropy
        i. first order notion (hartley) - first principles
            1. what function satisfies these conditions
                a. twice symbols ⇒ twice information, power of combinations
                b. $f(x^{\alpha}) = \alpha f(x)$ for all $x, \alpha > 0$
                c. logarithm
            2. $H = \beta \log s$ for s symbols
            3. assume equal probability of each symbol
        ii. second order notion (shannon) - probability theory
            1. not all symbols have the same probability
            2. compute expected value
            3. $H(x) = -\sum_{x} p(x) \log p(x)$ (show equivalence to hartley)
    c. entropy variants - these are just definitions based on statistical thinking
        i. joint entropy
            1. $H(x,y) = -\sum_{x,y} p(x,y) \log p(x,y)$
            2. over discrete sets this is the same as entropy
        ii. conditional entropy
            1. entropy of two variables conditioned on one
            2. $H(x|y) = -\sum_{x,y} p(x,y) \log [p(x,y) / p(y)]$
            3. $p(y)$ is the marginal probability
            4. $H(x|y) = H(x,y) - H(y)$
        iii. mutual information (transinformation)
            1. $I(x;y) = -\sum_{x,y} p(x,y) \log [p(x,y) / p(x)\, p(y)]$
            2. $I(x;y) = I(y;x) = H(x) + H(y) - H(x,y)$
            3. sum of entropies of the two marginals minus joint entropy
            4. important in shannon's cannon of work
        iv. information gain (kullback-leibler divergence)
            1. gain in information (between distributions)
            2. decrease in entropy / increase in order
            3. $D(p(X)|q(X)) = -\sum_{x} p(x) \log [p(x) / q(x)]$
            4. not commutative
            5. often computed by ad hoc means

4. **look at code**
    a. overview
        i. ruby with functional flair
        ii. attempt to easily see computational complexity
    b. basic computation
        i. /src/information_theory.rb
            1. almost one-liners
            2. $O(n)$
        ii. /src/symbols.rb
            1. convenience hash for demo
    c. examples
        i. rng.rb
            1. look at /src/randu.rb
            2. Random::srand is better than RANDU, but not great
            3. measure of disorder / order
        ii. hash.rb
            1. how well hashes work dicing up symbols
            2. similar hashing, not similar cryptographically
        iii. text_prob.rb
            1. just shows the probabilities of letters
        iv. text_ent.rb
            1. compare total entropies
            2. discuss different distributions
        v. waves.rb
            1. look at /src/wave_generator.rb
            2. symbol distribution
        vi. mutual.rb
            1. look at the code
            2. work the example on board - mutual information
            3. arrive at $1/4\ log\ 1/2\ +\ 3/4\ log\ 3/2$
        vii. classifier.rb
            1. preface with motivations regarding idea
                a. information gain
                b. bins
                c. recursion
            2. brief look at /src/classifier.rb methods
            3. observation of computational complexity
                a. $O(m \cdot n)$ for m attributes with n data elements
                b. $\sim m$ operations (recursively)
                c. $O(m^2 \cdot n)$ for full recursion
                d. in practice $m \ll n$
            4. run examples code
            5. discuss implementation issues
                a. trees structures
                b. continuous variables