CAS CS 365
Lec 9

Naive Bayesian(cont.) && confidence interval

1. Naive Bayesian → alphabet (cont.)
    a. Prior

    $$p(x) \propto \exp\left\{\frac{1}{2}\sum_{i\neq j}\beta_{ij}\left(x_i x_j + (1-x_i)(1-x_j)\right)\right\}$$

    b. Likelihood → try to fit the data well

    $$p(y|x) = \prod_{i=1}^{n} f(y_i|1)^{x_i} f(y_i|0)^{1-x_i}.$$

2. Exact Map Estimation for Binary images
    a. Our map inference becomes equivalent to minimizing

    $$\sum_{i=1}^{n} x_i \max(0, -\lambda_i) + \sum_{i=1}^{n} \max(0, \lambda_i)(1-x_i) + \frac{1}{2}\sum_{i\sim j}\beta_{ij}(x_i - x_j)^2,$$

    where $\lambda_i = \frac{f(y_i|1)}{f(y_i|0)}$.

    b. We have n * m binary matrix
    c. We impose a grid structure
    d. WE call two neighboring nodes bad if they have different values. We pay K units for each such pair
    e. We are allowed to flip the value of any node, but we have to pay R units
    f. The total cost is the sum of these two terms. How do we find the best assignment of values to nodes?
        i. R(# pixels flipped) + k(#disagreements)
    g. Max flow problem!
        i. Source s, sink t
        ii. Arc of capacity R from s to each node u with value 0
        iii. Arc of capacity R from each u node with value 1 to sink t
        iv. Directed arcs from each node u to its neighbors with capacity K
3. Convolution
    a. Suppose X and Y are independent, with known distributions, and Z = X + Y. What is distribution of Z?

i. Case 1 (discrete): the pmf is given by

$$p_Z(z) = \sum_q p_X(q) p_Y(z-q)$$

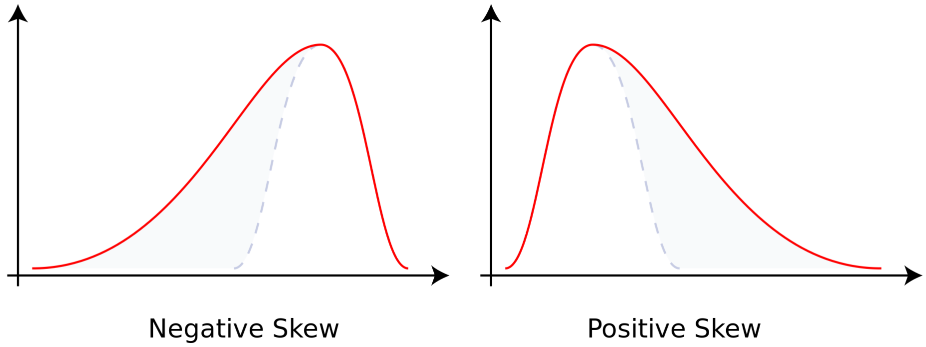
ii. Case 2 (continuous): the pdf is given by

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(q) f_Y(z-q)\,dq$$

4. Convolution of normal distribution
   a. The normal distribution satisfies a nice convolution identity:
   $$\text{if } X_1 \sim \mathcal{N}\left(\mu_1, \sigma_1^2\right), X_2 \sim \mathcal{N}\left(\mu_2, \sigma_2^2\right), \text{ then } X_1 + X_2 \sim \mathcal{N}\left(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2\right)$$
   b. Inductively we obtain the following fact:
   $$\text{if } X_1 \sim \mathcal{N}\left(\mu_1, \sigma_1^2\right), \ldots, X_n \sim \mathcal{N}\left(\mu_n, \sigma_n^2\right), \text{ then } \sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$
   c. If all Xi ~ N(μ, o^2), then by standardizing (setting mean to 0 and std to 1), we obtain
   $$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \sim \mathcal{N}(0, 1)$$
5. Moments
   a. We define the k-th moment of a random variable as follows:
   $$M_k(X) = \mathbb{E}\left[X^k\right]$$
      i. When k = 0, M(x) = 1
      ii. When k = 1, M(x) = E[x]
      iii. When k = 2, M(x) = Var(x) + M_{k-1}(x)^2
   b. We define the k-th central moment of a random variable as
   $$\mu_k(X) = \mathbb{E}\left[(X - \mathbb{E}X)^k\right]$$
      i. Therefore, μ_2(X) = Var(X)
6. Standardized moments of a RV
   a. Given a variable X with mean μ, and variance o^2, we define the standardized RV
   $$Z = \frac{X-\mu}{\sigma}$$
   b. The standardized moments of X are defined as the k-th moments of Z
   $$\bar{\mu}_k(X) = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^k\right]$$
7. Skewness and Kurtosis

a. Skewness → measures the symmetry of the distribution about mean:

$$\bar{\mu}_3(X) = \frac{\mu_3(X)}{\sigma^3}$$

i. 

ii. Negative skew: when left tail is longer (mean < median)

iii. Positive skew: when right tail is longer (mean > median)

b. Kurtosis: $$\bar{\mu}_4(X) = \frac{\mu_4(X)}{\sigma^4}$$

8. Applications: confidence intervals

a. Suppose we have 100 points from an unknown distribution, for which we know that the true population mean is 500, and the standard deviation is 80

i. Find probability that the sample mean will be inside the interval (490, 510)

1. CLT:
S_n - nμ
—--------- ~ N(0,1)
o*sqrt(n)
= S_n - nμ ~ N(0, o^2 * n)
= S_n - nμ
—----------- ~ N(0, o^2/n)
n
= S_n/n ~ N(μ, o^2/n)

2. By the CLT we know that the sample mean converges to $\bar{X}_n \to N\left(500, \left(\frac{80}{\sqrt{100}}\right)^2\right)$ in distribution. Therefore,

$$\Pr(490 \leq \bar{X}_n \leq 510) = \Pr\left(\frac{490-500}{\frac{80}{\sqrt{n}}} \leq \bar{Z}_n \leq \frac{510-500}{\frac{80}{\sqrt{n}}}\right) = \Phi(1.25) - \Phi(-1.25) = 0.789$$

3. Here, $$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x^2}{2}}\, dx$$ is the usual cdf

ii. Find an interval such that 95% of the sample average is covered

1. The mass on each side of the tails is at most (1-0.95)/2 = 0.025
2. Let's find the value z for which Φ(z)=1-0.025=0.975
3. Using calculator, the upper z value is 1.96, lower z value is -1.96
4. $$\sqrt{100}\frac{z-500}{80} = \pm 1.96 \Rightarrow z_{low} = 484.32,\ z_{up} = 515.68$$
5. In other words, the interval of 95% confidence is

$$\Pr\left(484.32\ \le \bar{X}_n \le 515.68\right) = 0.95$$