Computer Science 181

Joshua Lee
Homework 3
March 15, 2013

Saying that the maximum distance between any two dimensions is at most epsilon is equivalent to saying that the difference between all dimensions must be less than epsilon. Since we know $\boldsymbol{x}$ is placed in the center and $\varepsilon \in (0, 1/2)$, we know that all $\boldsymbol{y}$'s must fall into the range $(1/2 - \varepsilon, 1/2 + \varepsilon)$. Since $\boldsymbol{y}$'s are placed uniform randomly, the probability is proportional to length:

$$p(\max_m |x_m - y_m| \leq \varepsilon) = \prod_{m \in M} p(|x_m - y_m| \leq \varepsilon) = (2\varepsilon)^M$$

The window you want the dimensions of $\boldsymbol{y}$'s is always at most 2ε. This is maximal and true for all dimensions when $\boldsymbol{x}$ is placed at the center of the hyper cube. Now that $\boldsymbol{x}$ is being placied randomly, if some of the dimensions of $\boldsymbol{x}$ get close to the edge of the hypercube (i.e. close to 0 or 1), then the window the corresponding $\boldsymbol{y}$ dimension can fall into is smaller than 2ε, therefore the probability is less.

We have already shown that $p(\max_m |x_m - y_m| \leq \varepsilon) \leq (2\varepsilon)^M)$. Note that $||\boldsymbol{x} - \boldsymbol{y}|| \geq \max_m |x_m - y_m|$. This is obvious because the square root of the sum of the squared differences between all dimensions is clearly greater than the maximum difference between any one dimension. Therefore, we know that $||\boldsymbol{x} - \boldsymbol{y}||$ must have a SMALLER probability than $\max_m |x_m - y_m|$ of being less than ε, precisely because $||\boldsymbol{x} - \boldsymbol{y}|| \geq \max_m |x_m - y_m|$. Therefore, $p(||\boldsymbol{x} - \boldsymbol{y}|| \leq \epsilon) \leq (2\varepsilon)^M$

Let $p = (2\varepsilon)^M$. Using our result from (c), the probability that a single point will be greater than ε away from $\boldsymbol{x}$ at least $(1 - p)$. So, the probability all N points are greater than ε away from $\boldsymbol{x}$ at least $(1 - p)^N$. Therefore the probability that at least 1 point is at most ε away from $\boldsymbol{x}$ is $1 - (1 - p)^N$. So we have that $1 - (1 - p)^N \geq 1 - \delta$.

Rearranging the inequality from (d), we have that:

$$N \geq \frac{\ln(\delta)}{\ln(1 - (2\varepsilon)^M)}$$

This means that to have at least probability $(1 - \delta)$ of having 2 points be less than ε apart, we need at least N points, which increases as M increases! Thus, as we have higher dimension data, we need more data to have the same probability of datum to be close to each other. This is the mathematical formulation of what we termed the "Curse of Dimensionality" in lecture.

Maximum Likelihood (ML)

$\theta_{\mathrm{ML}} = \arg\max\limits_{\theta} Pr(\mathcal{D}|\theta)$ (e.g. $\frac{N_1}{N}$ in the case of Bernoulli)

Maximum a posteriori (MAP)

$\theta_{\mathrm{MAP}} = \arg\max\limits_{\theta} Pr(\mathcal{D}|\theta)Pr(\Theta = \theta)$ where $Pr(\Theta = \theta)$ is our prior.
(e.g. $\frac{\alpha+N_1-1}{\alpha+\beta+N-2}$ in the case of Bernoulli).

Full Bayesian (FB)

$Pr(\boldsymbol{x}|\mathcal{D}) = \int_{\boldsymbol{\Theta}} Pr(\boldsymbol{x}|\boldsymbol{\theta})Pr(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}$

Using each method's respective θ's, we can then calculate $Pr(\boldsymbol{x}|\mathcal{D})$ using the model distribution we choose. For example, in the case we choose to model using a Bernoulli distribution, we have that $Pr(\boldsymbol{x}|\mathcal{D}) = \theta$.

Let $Pr(w_4|\theta)$ be the probability the Harvard football team wins the fourth game of the 2012-13 season, given the parameters θ. We use the Bernoulli distribution as our model distribution.

2011-12 record: 9 wins - 1 loss
2012-13 record: 3 wins - 0 losses (first three games)

Maximum Likelihood (ML)

For ML, we do not use a prior, so we ignore the data from the 2011 - 12 season.
Using the data from the 2012 - 2013 season:

$$N = 3, \ N_1 = 3, \ N_0 = 0, \ \theta_{\text{ML}} = \frac{N_1}{N} = 1$$

Therefore, $Pr(w_4|\theta_{\text{ML}}) = \theta_{\text{ML}}^{w_4}(1 - \theta_{\text{ML}})^{(1-w_4)} = 1^{w_4} = 1$

Solution: $Pr(w_4|\theta_{\text{ML}}) = 1$

Maximum a posteriori (MAP)

For MAP, we use the prior distribution:

$$Pr(\Theta = \theta) = \text{Beta}(\Theta = \theta|\alpha, \beta) = \frac{1}{Z}\theta^{\alpha-1}(1 - \theta)^{\beta-1}$$

where $Z = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ (normalizing constant) and $\alpha = 9, \beta = 1$ (using the 2011-12 season data).

Now we want to update our prior based on the data, giving us:

$$Pr(\Theta = \theta|\mathcal{D}) = \text{Beta}(\Theta = \theta|9 + N_1, 1 + N - N_1) = \text{Beta}(\Theta = \theta|12, 1)$$

since Beta is the conjugate prior of the Bernoulli and we have $N_1 = 3, N = 3$. So:

$$\theta_{\text{MAP}} = \frac{\alpha + N_1 - 1}{\alpha + \beta + N - 2} = \frac{9 + 3 - 1}{9 + 1 + 3 - 2} = 1$$

Therefore, $Pr(w_4|\theta_{\text{MAP}}) = \theta_{\text{MAP}}^{w_4}(1 - \theta_{\text{MAP}})^{(1-w_4)} = 1^{w_4} = 1$

Solution: $Pr(w_4|\theta_{\text{MAP}}) = 1$

Full Bayesian (FB)

Again we use the prior $\Pr(\boldsymbol{\theta}) = \text{Beta}(\boldsymbol{\theta}|9,1)$ and posterior $\Pr(\boldsymbol{\theta}|\mathcal{D}) = \text{Beta}(\boldsymbol{\theta}|12,1)$ we calculated above.

Using the FB method where we now consider the distribution over all possible parameter values, we have:

$$\theta_{\text{BF}} = \int_0^1 Pr(w_4|\theta)Pr(\theta|\mathcal{D})d\theta = \frac{\alpha + N_1}{\alpha + \beta + N} = \frac{9+3}{9+1+3} = \frac{12}{13}$$

Therefore, $Pr(w_4|\theta_{\text{BF}}) = \theta_{\text{BF}}^{w_4}(1 - \theta_{\text{BF}})^{(1-w_4)} = \frac{12}{13}^{w_4}\frac{1}{13}^{(1-w_4)}$

Solution: $Pr(w_4|\theta_{\text{BF}}) = \frac{12}{13}^{w_4}\frac{1}{13}^{(1-w_4)}$