

Assignment 2

Instructor: Murat Dunder

Due date: 2/23/2017

Part A: Dirichlet-Multinomial Conjugate Model for Classification

First, read the material “Bayesian inference, entropy, and the multinomial distribution” by Thomas Minka until the end of page 3. This material is available at:

<http://research.microsoft.com/en-us/um/people/minka/papers/minka-multinomial.pdf> until the end of page 3.

A.1. (5 points) Prove equation 21, which is the posterior predictive distribution for the Multinomial-Dirichlet model.

A.2. (5 points) Prove $p(\mathbf{Y}|\mathbf{X}, \alpha)$ in equation 28.

A.3. (15 points) Write a Matlab script that will implement equation 28 for a K-class classification problem.

A.4. (15 points) Test your script from A.3. on the PubMed data set provided. Train with the training data and test on the test data. Compute the mean F1 score of your classifier. Tune α to optimize classifier performance on the test data.

Part B: Normal-Normal-Inverted Wishart Conjugate Model for Classification

Data Model: $x \sim N(\mu, \Sigma)$, $\mu \sim N(\mu_0, \kappa^{-1}\Sigma)$, $\Sigma \sim W^{-1}(\Sigma_0, m)$.

Given a data set $D = \{x_1, \dots, x_n\}$, $x_i \in R^d$, generated by this model, let the sample mean be $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ and the sample covariance matrix be $S = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T$.

B.1. (5 points) Show that $\bar{x} \sim N(\mu, n^{-1}\Sigma)$

B.2. (5 points) Show that $p(\mu, \Sigma | \bar{x}, S) = N\left(\frac{n\bar{x} + \kappa\mu_0}{n + \kappa}, \frac{\Sigma}{n + \kappa}\right) \times W^{-1}\left(\Sigma_0 + (n - 1)S + \frac{n\kappa}{n + \kappa}(\bar{x} - \mu_0)(\bar{x} - \mu_0)^T, n + m\right)$ (hint: $(n - 1)S \sim W((\Sigma, n - 1))$)

B.3. (5 points) Show that $p(x^* | \bar{x}, S) = st u - t\left(\frac{\kappa\mu_0 + n\bar{x}}{n + \kappa}, \frac{n + \kappa + 1}{(n + \kappa)(n + m + 1 - d)} \left[\Sigma_0 + (n - 1)S + \frac{n\kappa}{\kappa + n}(\mu_0 - \bar{x})(\mu_0 - \bar{x})^T\right], n + m + 1 - d\right)$ (hint: $|A + xx^T| = |A|(1 + x^T A^{-1}x)$)

B.4. (20 points) Write a Matlab script that will implement an MAP classifier using the posterior predictive distribution in B.3. for a K-class classification problem.

B.5. (25 points) Test your script for B.4. on the PudMed data set provided. Note that PubMed data is in a bag-of-words format. Before you can use this data for this task you need to apply preprocessing techniques. Experiment with PCA, SVD, and NNMF to see how different data reduction techniques affect the results. Train with the training data and test on the test data. Compute the mean F1 score of your classifier. Tune κ, μ_0, Σ_0 and m to optimize classifier performance on the test data.

Bonus (5 points): Which model did better and why?

Derivations of B.2. and B.3. are not trivial. I have assigned only 5 points for each so that you do not get penalized much in case you choose to skip derivations. For this question you will need the pdf's of multivariate Student-t, Wishart, and Inverse-Wishart distributions which you can all get from Wikipedia.