MATH 156S HOMEWORK 5

Due May. 5

Exercise 1. According to Exercise 2.7.5 we can apply multinomial naive Bayes for the image classification of MNIST dataset, since the features there (pixel values) are nonnegative.
(Ref: Math156_classification_MNIST.ipynb and Math156_20NewsGroups.ipynb)

(i) Reproduce Figure 16 for classifying digits '4' and '7' using multinomial naive Bayes. Compare the performance with logistic regression and probit regression.

(ii) Reproduce Figure 19 for classifying digites 0 through 4 using multinomial naive Bayes. Compare the performance with multiclass logistic regression.

Exercise 2. In this exercise, we investigate the 'robustness' of multinomial naive Bayes (MNB) classifier on the 20Newsgroups dataset. (Ref: Math156_20NewsGroups.ipynb)

(i) Modify the data preprocessing function "sample_multiclass_20NEWS" in the Jupyter notebook so that it has additional parameters 'noise_ratio' and 'noise_intensity': If 'noise_ratio' is $\varepsilon \in [0,1]$, and 'noise_intensity' is $\delta > 0$, then then it will add independent Uniform($[0,1]$) variables times δ to ε density of randomly chosen coordinates.

(ii) Choose two categories in the 20Newsgroups dataset and compare the performance of binary classification using MNB with varying noise parameters $(\varepsilon, \delta) \in \{0.1, 0.5, 0.9\} \times \{0.001, 0.01, 0.1\}$. (Use tf-idf vectorizer.) Do you think if MNB is robust against noise on the 20Newsgroups dataset? If so, is it more robust against noise ratio (ε) or noise intensity (δ)? Otherwise, can you say why?

(iii) Perform similar experiments as in **(ii)** for multiclass logistic regression.

Exercise 3. Let ℓ_{GNB} denote the Gaussian naive Bayes loss function defined in (173). Denote the solution of (173) as $\hat{\mathbf{W}} = (\hat{\mathbf{q}}_i, \hat{\mu}_{ij}, \hat{\sigma}_{ij})$.

(i) Argue that the optimal prior PMF $\hat{\mathbf{q}} := [\hat{q}_1, \dots, \hat{q}_\kappa]$ solves the following optimization problem

$$\hat{\mathbf{q}} = \underset{\substack{\mathbf{q}=[q_1,\dots,q_\kappa] \\ \text{PMF on } \{1,\dots,\kappa\}}}{\arg\max} \sum_{i=1}^{\kappa} \left(\sum_{s=1}^{N} \mathbf{1}(y_s = i) \right) \log q_i. \tag{1}$$

Conclude that $\hat{q}_i = \frac{1}{N} \sum_{s=1}^{N} \mathbf{1}(y_s = i)$ (Hint: Use Exercise 2.7.2).

(ii) Show that

$$\frac{\partial \ell_{GNB}(\mathbf{W})}{\partial \mu_{ij}} = \sum_{s=1}^{N} \mathbf{1}(y_s = i) \frac{(\phi_j(\mathbf{x}_s) - \mu_{ij})}{\sigma_{ij}^2}. \tag{2}$$

Deduce that $\hat{\mu}_{ij}$ equals the following 'sample mean of the jth feature in class i':

$$\hat{\mu}_{ij} = \frac{\sum_{s=1}^{N} \mathbf{1}(y_s = i) \phi_j(\mathbf{x}_s)}{\sum_{s=1}^{N} \mathbf{1}(y_s = i)}. \tag{3}$$

(Compare this with $\hat{q}_{ij}$ in Prop. 2.7.1.)

(iii) Show that

$$\frac{\partial \ell_{GNB}(\mathbf{W})}{\partial \sigma_{ij}} = -\sigma_{ij}^{-1} \left(\sum_{s=1}^{N} \mathbf{1}(y_s = i) \right) + \sigma_{ij}^{-3} \sum_{s=1}^{N} \mathbf{1}(y_s = i)(\phi_j(\mathbf{x}_s) - \mu_{ij})^2. \tag{4}$$

Deduce that $\hat{\sigma}_{ij}$ equals the following 'variance of the class-i empirical distribution':

$$\hat{\sigma}_{ij}^2 = \frac{\sum_{s=1}^{N} \mathbf{1}(y_s = i)(\phi_j(\mathbf{x}_s) - \hat{\mu}_{ij})^2}{\sum_{s=1}^{N} \mathbf{1}(y_s = i)}, \tag{5}$$

where $\hat{\mu}_{ij}$ is given in **(ii)**.

Exercise 4. Derive the total loss function in (184) using the square loss in (185) by using the maximum likelihood framework in (189). You may assume that the true output $\mathbf{y} \in \mathbb{R}^\kappa$ is generated by a k-diemnsional multivariate Normal distribution $N(\hat{\mathbf{y}}(\mathbf{x}; \mathbf{w}), \sigma^2 I)$, where I is the $\kappa \times \kappa$ identity matrix. (See Example 3.2.2.)