ASSIGNMENT 3

Deadline: 11:59 pm, March 31, 2024

Submit via Blackboard with VeriGuide receipt.

Please follow the course policy and the school's academic honesty policy.

1. Deterministic noise depends on $\mathcal{H}$, as some models approximate f better than others (referring to week 7's slide page 8).

   (1) Assume $\mathcal{H}$ is fixed and we increase the complexity of f. Will deterministic noise in general go up or down? Is there a higher or lower tendency to overfit? Please provide a detailed analysis.

   Hint: Compare the complexity between $\mathcal{H}$ and f.

   (2) Assume f is fixed and we decrease the complexity of $\mathcal{H}$. Will deterministic noise in general go up or down? Is there a higher or lower tendency to overfit? Please provide a detailed analysis.

   Hint: There is a race between two factors that affect overfitting in opposite ways, but one wins.

2. In this problem, we will focus on the regularization (referring to week 7's slide page 19 and page 26).

   Supposed that the regularized weight $\mathbf{w}_{\text{reg}}$ is a solution to:

$$\begin{aligned} &\text{min: } E_{\text{in}}(\mathbf{w}) = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{w}^\intercal \mathbf{x}_n - y_n)^2 \\ &\text{subject to: } \mathbf{w}^\intercal \Gamma^\intercal \Gamma \mathbf{w} \leq C \end{aligned}$$

   where Γ is a matrix. If $\mathbf{w}_{\text{lin}}^\intercal \Gamma^\intercal \Gamma \mathbf{w}_{\text{lin}} \leq C$, where $\mathbf{w}_{\text{lin}}$ is the linear regression solution, then what is the regularized weight $\mathbf{w}_{\text{reg}}$?

   [a] $\mathbf{w}_{\text{reg}} = \mathbf{w}_{\text{lin}}$

   [b] $\mathbf{w}_{\text{reg}} = C\mathbf{w}_{\text{lin}}$

   [c] $\mathbf{w}_{\text{reg}} = \Gamma\mathbf{w}_{\text{lin}}$

   [d] $\mathbf{w}_{\text{reg}} = C\Gamma\mathbf{w}_{\text{lin}}$

   [e] $\mathbf{w}_{\text{reg}} = \Gamma^\intercal\Gamma\mathbf{w}_{\text{lin}}$

   [f] None of the above

   Please select **one correct option** from the above options and briefly discuss your reason.

   Hint: $w_{lin} = (Z^\intercal Z)^{-1} Z^\intercal y$ and $w_{reg} = (Z^\intercal Z + \lambda * I)^{-1} Z^\intercal y$.

3. Supposed that 60% of mobile phones are model A and 40% are model B. Due to the high similarity of the mobile phones, the probability of a tech enthusiast correctly identifying the phone model is 90%. If a tech enthusiast claims to have just seen a mobile phone of model A, what is the probability that the mobile phone is indeed model A?

   Hint: Bayes rule (referring to week 8's slide page 8): $p(y|x) = \frac{p(x|y)p(y)}{p(x)}$.

4. The Poisson distribution is one of the most popular distributions in statistics. You can learn more about it on Wikipedia. The probability density function (PDF) of the Poisson distribution is given as:

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Given a sample of n measured values $x_i \in \{0, 1, ...\}$, for $i = 1, ..., n$, we wish to estimate the value of the parameter λ of the Poisson population from which the sample was drawn. Please derive the maximum likelihood estimate $\hat{\lambda}$.

Hint: You could calculate the log likelihood function first and utilize the derivative condition for maximum value to calculate the result.

5. Gaussian Discriminant Analysis (GDA) can be utilized for binary classification tasks (referring to week 8's slide page 27). Considering the datasets in Figure 1, please identify which dataset is more suitable for GDA and briefly discuss your reason.

(a) Dataset A (b) Dataset B

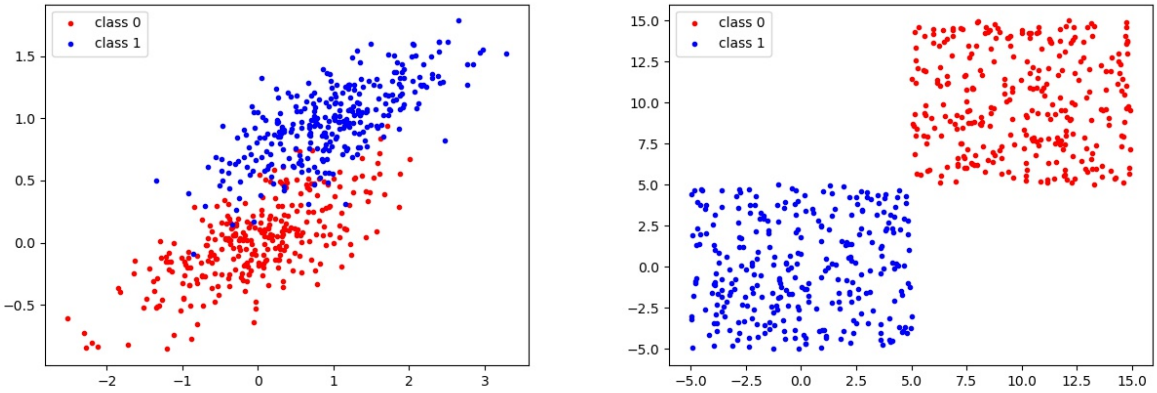

Figure 1

6. Given the GDA model's discriminant function

$$\log p(y = c \mid \boldsymbol{x}, \boldsymbol{\theta}) = \log \pi_c - \frac{1}{2} \log |2\pi \boldsymbol{\Sigma}_c| - \frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu}_c)^\top \boldsymbol{\Sigma}_c^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_c) + \text{ const} . \quad (1)$$

Please describe why the decision boundary between any two classes will be a quadrative function of x. In addition, in which case the LDA (figure 2 (right), referring to week 8's slide page 29) will be a special case of GDA and please explain that according to the decision boundaries that can be derived from the equation 1.

7. Python Programming Task

(1) Generate a dataset containing 6000 data points from three different Gaussian distributions.

For Gaussian distribution 1: means=$[0, 0]^T$, variance=$diag([1, 2]^T)$.

For Gaussian distribution 2: means=$[4, 0]^T$, variance=$diag([2, 2]^T)$.

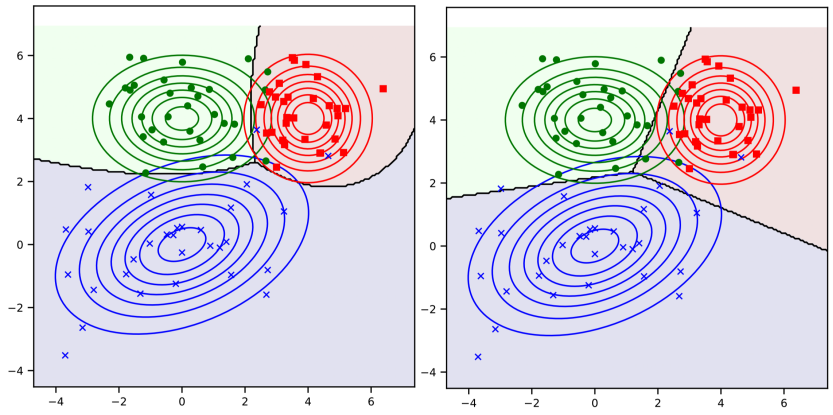

Figure 2

For Gaussian distribution 3: means=$[2, 0]^T$, variance=$diag([2, 1]^T)$.

The intensity ratio of the three Gaussian distributions is 3:2:2. After generating, visualize the results and mark data from different distributions with different colored points.

(2) Fit a Gaussian Mixture Model with three components to the dataset generated in the previous step. Classify each point in the dataset with the fitted model. Provide the precision and recall of the fitting results for each class and plot the prediction result in different colors. Compare the figure you get here with that in the first step. Note that for simplicity, we assume the covariance matrices of each Gaussian distribution are diagonal. There are different prediction-groundtruth matching choices. Use the one can maximize the precision!

(3) In the second step, remove the assumption that the covariance matrices of each Gaussian distribution are diagonal, refit, and predict the category of each point. Provide the new precision and recall. Plot the prediction results in a figure and compare it with that in the second step.

(4) In the third step, repeat the fitting with component numbers of 2, 3, 10, and 100. Plot the prediction results in a figure. Discuss how the hyperparameter "number of components" affects the fitting results.

Hint: You can refer to some python packages from numpy, matplotlib and sklearn. For example, we can directly use **from sklearn.mixture import GaussianMixture** and **np.random.multivariate_normal()**.

Note: Please submit your python implementation (e.g., Q7.py, Q7.ipynb) along with the report. Here is a submission example via Blackboard for your reference: **report.pdf**, **Q7.py**, **veriguide_receipt.pdf**. Please note that there are no restrict limitations regarding the submission format.

*** END ***