midterm

November 3, 2021

1 STP598 Machine Learning & Deep Learning

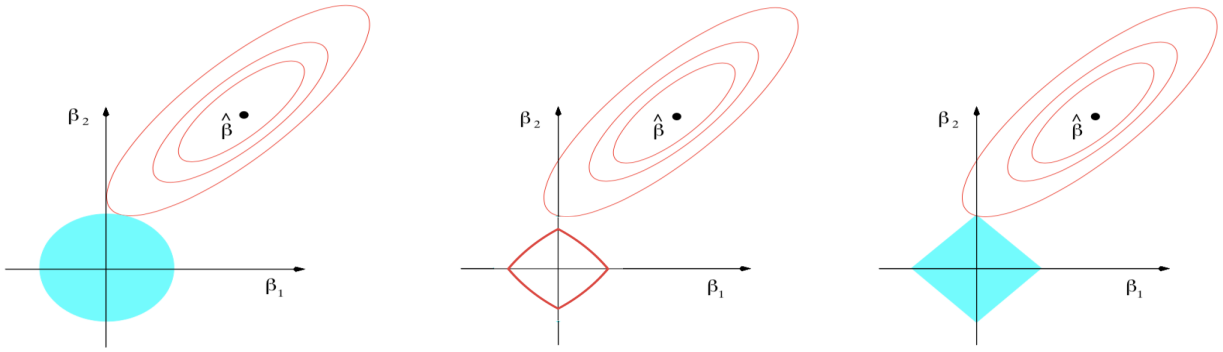
1.1 Midterm Exam (Take-home)

1.1.1 Due 11:59pm Wed Nov. 10, 2021 on Canvas

1.1.2 name, id

1.2 Question 1

- In multiple linear regression, we have residual vector defined as $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$. Prove that it is perpendicular to the column space of $\mathbf{X}$, i.e. $\mathbf{X}^T\mathbf{e} = \mathbf{0}$.
- Give each of the following figure the correct description: 'elastic net', 'lasso' or 'ridge regression'?



1.3 Question 2

Probit regression is a binary classification model alternative to logistic regression. The link function is probit function (inverse CDF of standard normal Φ^{-1}) instead of logit function, i.e.

$$\Phi^{-1}(\Pr(Y = 1|X)) = X\beta$$

- Write down the log-likelihood function of $\{x_i, y_i\}_{i=1}^n$ and answer briefly how you can find the solution $\hat{\beta}$.
- Fit **iris** data (only consider setosa vs non-setosa, i.e. convert labels 2 to 1) with *logistic regression* and *probit regression* respectively and plot their decision boundaries side by side. For

illustrative purpose, you can use the first predictors. Hint: probit regression is not implemented in `scikit-learn` but has been implememted in `statsmodels`. Consider statsmodels.discrete.discrete_model.Probit

1.4 Question 3

Compare impurity measures for splitting nodes in trees.

- Fill in the blanks of the table to compute Gini index, Shannon entropy and misclassification error.

	Class 1	Class 2	Class 3	$\hat{p}_1$	$\hat{p}_2$	$\hat{p}_3$	Gini	Entropy	Error
$\mathcal{A}$	3	3	4						
$\mathcal{A}_L$	1	0	3						
$\mathcal{A}_R$	2	3	1						

- Compute the impurity reductions for the **three** measures.

1.5 Question 4

Gaussian process is a flexible tool for modeling nonlinear functional relationship. Given data $\{x_i, y_i\}_{i=1}^n$, we assume the following model:

$$y_i = f(x_i) + \epsilon_i, \quad \epsilon_i \overset{iid}{\sim} N(0, \sigma_\epsilon^2)$$
$$f \sim \mathcal{GP}(0, \mathcal{C})$$

- Given a new location x_*, predict $\hat{y} = f(x_*)$ and give the uncertainty estimate (credible interval).
- Simulate a dataset of 1-d input x and output y, e.g. using $y = \sin(x) + .1 * N(0,1)$, for $n_1 = 10$ points. Use Gaussian process to fit such dataset. Predict x_* on a grid of 100 points over the defined domain ($[0, \pi]$ for example). Now increase the data to $n_2 = 50$ points (may contain n_1 points), repeat the same prediction. Plot the following on the same graph:
  - n_1 data points and n_2 data points with different colors (scatter plot)
  - posterior prediction lines based on n_1 and n_2 respectively with different colors (line plot).
  - posterior credible bands based on n_1 and n_2 respectively with different colors (fill_between)
- Compare the plots between two cases (n_1 vs n_2). What do you find?

1.6 Question 5

Use the same simulation dataset with $n_3 = 1000$ points. Fit neural network (nn) models.

- Fit a single hidden layer (e.g. with 100 units) nn and a multiple (e.g. 10 or more layers) hidden layer nn respectively. Compare the fitted curve. Did you observe overfitting?
- Do you have uncertainty estimates from nn models directly? If not, do you have any ideas to get uncertainty (interval) estimates?

1.7 Extra*

Please comment on this course. What suggestions do you have to improve this course?