DATA7703 Tutorial 11 Solutions

2021 Semester 2

1. **(a)** In the built-in approach, models are designed to be interpretable (e.g. linear regression). In the post hoc approach, models are analyzed for interpretability (e.g. permutation importance)

   (b) In a white-box method, everything about the model is needed (e.g. linear regression model weights). In a black-box method, only partial information about the model is needed (e.g. permutation importance).

   (c) A model-specific is designed for specific models only (e.g. linear regression model weights), while a model-agnostic method is designed for generic learning approaches (e.g. permutation importance).

2. False. In a linear model, the coefficients are generally poor indicators of the importances of the features. This is because the features may be on different scales, and the coefficient of a feature becomes larger when a larger unit is used, while the importance of the feature should remain unchanged.

3. A column for a permuted dataset should be a permutation of the column for the original dataset. D_1 does not satisfy this property, so it cannot be a permuted dataset.

4. A Gaussian process is a collection of random variables such that any finite subset of them follows a Gaussian distribution. Any multivariate Gaussian distribution is a Gaussian process.

5. **(a)** $Y_1 \sim N(0, \sigma_1^2)$, $Y_2 \sim N(0, \sigma_2^2)$.

   (b) In a multivariate normal distribution, the (i, j)th entry of the covariance matrix is the covariance of the ith and jth random variables. From this, we have $\mathrm{Var}(Y_1) = \sigma_1^2$, $\mathrm{Var}(Y_2) = \sigma_2^2$, $\mathrm{cov}(Y_1, Y_2) = \rho\sigma_1\sigma_2$. The correlation of Y_1 and Y_2 is $\frac{\mathrm{cov}(Y_1,Y_2)}{\sqrt{\mathrm{Var}(Y_1)\,\mathrm{Var}(Y_2)}} = \rho$.

   (c) The variance of Y_1 is σ_1^2.

   For the variance of Y_1 given $Y_2 = y_2$, note that $Y_1 \mid Y_2 = y_2 \sim N(\rho\sigma_1\sigma_2(\sigma_2^2)^{-1}y_2, \sigma_1^2 - (\rho\sigma_1\sigma_2)(\sigma_2^2)^{-1}(\rho\sigma_1\sigma_2))) = N(\frac{\rho\sigma_1}{\sigma_2}y_2, (1-\rho)\sigma_1^2)$. Thus the conditional variance of Y_1 is $(1-\rho)\sigma_1^2$.

   Hence the conditional variance is smaller than the unconditional variance.

   Intuitively, an observation on a correlated random variable reduces the variance of a random variable.