Q51-WilliamKennedy-300015367

William Kennedy

2023-12-10

1. When is ridge regression preferable to LASSO regression?

ChatGPT said regression wants to punish you for your mistakes but not completely get rid of you. To me this sounds accurate, but this feels like a layman description however it does capture the essence of Ridge regression. When the hyperparameter is small the penalty is small and vice versa. Although it doesn't describe that when the magnitude continues being large after being penalized that the parameter must be really important.

ChatGPT described LASSO as sending coefficients to zero if the algorithm deems them unnecessary. This is accurate, when the magnitude is large then the coefficient goes to 0. But this is again a layman expression, it captures the main idea of LASSO but doesn't give all the important details.

Most importantly, ChatGPT didn't answer the question. It did not tell us Ridge Regression is preferable over LASSO Regression. One possibility is when there are a large number of predictors, Ridge Regression would be preferable. For however you define "large" within the context of the data.

2. What is the naive assumption in the naive Bayes classifier?

ChatGPT is completely correct in describing predictors as being independent of each other. Though I do think that the example provided was not a very good one, to me Carla's love of dance and her hatred of Turks "World's Best Surgeon" sound quite independent to me. But maybe this is ChatGPT trying to be funny, which is even more concerning. . . . . .

3. A classifier is trained on a cancer dataset and achieves 96% accuracy on new observations. Why might this not be considered a good classifier?

It is correct about it being dangerous for missclassifications, 96% may be a good classification rate for vetting resumes. But when determining cancer rates, this is a huge and unnacceptable margin of error. Additionally, a model with a high accuracy is not necessarily robust

4. A regression model has low bias and high variance. How can it be improved?

When a model has a high variance the model has overfit the data, so the model is too 'loose' on the data. It is true that Ridge Regression does help to mitigate the bias-variance trade-off and reduce overfitting,but adding more observations may not necessarily fix this if there were a sufficient number of observations to begin with since the bias-variance trade-off is a result of the model being used. The metaphor is interesting but I don't it's completely descriptive of a model with low bias and high variance, yes it articulates to simplify the model but saying it gives every single strand a story does not directly connect to the idea of the model interpreting noise as useful

5. How is kNN different from k-means clustering? I like this description, I think it's quite correct. KNN is used in classification when there are too few observations with specific feature values. So it classifies a response based on the number of nearest neighbours, however the metaphor describes these neighbours as similar which isn't necessarily true. They're not the most similar neighbours, they're the nearest. Nearest does act as a proxy for similarity but I think this could make someone confused about the difference between KNN and k-means clustering. As for k-means clustering I think is a weird description, it doesn't really tell me what k-means clustering is. A better example might be "lets group together this lot of cars, by year and make" as our similarity measure.

I think this description does not do a good job of differentiating the two methods.

6. List 6 feature selection/dimension reduction methods

Yes, these are all feature selection and dimension reduction methods, though I don't walk away from this description with any meaningful understanding or really any understanding of what these methods do.