Linear Regression Quiz

1. Regression is used when the thing being predicted is a real number, and Classification is used when the thing being predicted is a category or label from a set of labels.

2. There can be 1, 2, or many number of features for the input data on a Linear Regression problem.

3. If we have p training instances each with q features, then to train the Linear Regression model we need to create a "design matrix". How many rows and columns will this have?

   p rows and q+1 columns.

4. In the design matrix, each row represents one training instance.

5. If we denote the i-th row of the design matrix $\underline{x}_i^T = (1, x_{i1}, x_{i2}, \ldots)$, and if the linear regression model has parameters $\underline{w} = (w_0, w_1, \ldots, w_q)$, what's the model's prediction for training instance j?

   $\underline{w}^T \underline{x}_j$

6. If Φ is the design matrix and $\underline{w}$ the linear regression model parameters, then $\Phi\underline{w}$ is the vector of predictions with the j-th component being the model's prediction for training instance j.

7. We can create an error function to assess how good the predictions are, commonly squared error:

   $$E(\underline{w}) = \sum_{i=1}^{n} (y_i - \underline{w}^T \underline{x}_i)^2$$

   error; sum over the training patterns; y_i: correct output from training data (true label); $\underline{w}^T \underline{x}_i$: model's predicted output

8. To find a way to minimise the error function we note that at the minimum, the partial derivatives of the error function with respect to the model parameters are zero.

9. In the probabilistic interpretation of the error function approach to finding the best model parameters **w**, which probabilistic right hand item relates to which error-function left hand one?

minimal error - maximal log likelihood
sum of squares - variance

10. Linear regression is sensitive to outliers!

11. Should outliers be removed from the training data, should they also be removed from the testing data? NO!

12. Should outliers be removed before or after the linear model is fit to the data? Before - linear preprocessing.

13. Good ways to increase the chances that a basic linear model works for your dataset:
- Visualise the data and remove outliers before fitting
- Fit the model then visualise the residuals against actual outputs, expecting no correlation

14. Multiple regression includes which of the following scenarios?
- Multiple attributes are used to predict multiple output values
- One attribute is used to predict multiple output values

15. How to do multiple regression: build separate regression models for each output variable

16. When we do basis expansion, we are transforming the input data into a new input feature space.

17. We can use categorical attributes with linear regression by assigning each category a new binary attribute.

18. RBF facts
- RBF have two parameters
- a good way to pick RBF centres is to select a few of the data points
- if you have too many RBFs, you run the risk of overfitting