Q1. You are given a qualitative attribute, X_{colour}, taking 7 possible values, i.e.

$$X_{colour} \in \{colour_1, \dots, colour_K\}$$

You would like to transform it into a set of binary dummy variables to be added instead of X_{colour}
- to the following linear regression model

$$F(X_{quantitative}, \lambda) = [1, x^T_{quantitative}]\lambda$$

· where $X_{quantitative} \in \mathbb{R}^d$ with $d = 9$.

What is the number of degrees of freedom of the model after you have added the new dummy variables?

Q2. Let D be a random data set generated by the true process

$$Y \sim f(X) + \epsilon$$

where ϵ is random.

Let $\hat{f}(x_0, D)$ the prediction for new input, x_0, made by a model trained on D.

a. The bias of $\hat{f}(X, D)$ at x_0 depends on x_0.

c. The bias of $\hat{f}(X, D)$ at x_0 may depend on the degrees of freedom of the model class.

f. In general, squared bias is low for models that overfit the data.

h. The bias of $\hat{f}(X, D)$ at x_0 is $E_D(\hat{f}(x_0)) - f(x_0)$

Q3. Suppose you train two prediction algorithms, A0 and A1, and are looking for evidence that

- A1 produces a better prediction rule than A0.

~~and you apply the two prediction rules to the same dataset.~~

Your null hypothesis is that

- A1 is not better than A0

and you apply the two prediction rules to the same dataset. The results are summarized in the following contingency table:

	A1	
A0	correct	wrong
correct	1890	43
wrong	59	231

- There is evidence that the prediction rule produced by A1 is better.
  - the evidence is not statistically significant at level 5%
  - the evidence is not highly statistically significant at level 1%.

You can use R to answer this question.

Q4. You are training a linear regression learning machine
$F(X,\lambda) = \lambda_0 + \lambda_1 X$ on the following data set.

$\mathcal{D} = \{(x_n, y_n) \in \mathbb{R}^2\}_{n=1}^{N} = \{(0,1), (1,1.5), (2,-0.5), (3,-2)\}$
Compute the value of the ℓ_2-penalised objective

$$J(\mathcal{D},\lambda) = \rho\|\lambda\|^2 + \sum_{n=1}^{N}(y_n - F(x_n,\lambda))^2, \quad \rho = 0.1$$

for $\lambda = [1, -1]^T$.

Q5. You test a classification algorithm on a given test data set with labels $Y \in \{yes, no\}$ and obtain

	True label	
Prediction	"yes"	"no"
"yes"	11	5
"no"	7	9

TP	FP
FN	TN

Compute the error rate of the algorithm.

The total number of observations: $11+5+7+9=32$

- Prediction's error rate $= \dfrac{FN+TN}{\text{Total observations}}$

$$= \frac{7+9}{32} = \frac{16}{32} = \frac{1}{2} = 0.5\,\%$$

- True label's error rate $= \dfrac{5+9}{32} = \dfrac{14}{32} = 0.4375\,\%$

$\therefore$ The error rate of the algorithm $= 0.5+0.4375 = 0.9375\,\%$