

Regressão Logística

Nome da Atividade		
Regressão Logística		
1	Texto de Apresentação	<p>Problema 1</p> <p>Vamos fazer um exercício de Regressão Logística, trabalhando com a base de dados do Titanic.</p> <p>Utilizaremos os arquivos disponibilizados no Moodle que foram preparados para esta atividade. Embora os nomes sejam os mesmos, estes arquivos do Moodle são diferentes (em conteúdo) daqueles encontrados no site Kaggle.</p> <p>A base de dados (treinamento) é uma coleção de dados sobre alguns dos passageiros (916 para ser preciso) e o objetivo é prever a sobrevivência (1 se o passageiro sobreviveu ou 0 caso contrário), baseado em algumas características tais como classe de serviço, sexo, idade, etc. A base de dados de teste tem a mesma estrutura da base de treinamento, mas a variável survived não contém nenhum valor.</p> <p>Depois de construir seu modelo, você vai aplicá-lo ao este conjunto de dados de teste e prever o valor da variável survived.</p> <p>Como se pode ver, utilizaremos tanto variáveis categóricas como contínuas.</p> <p>Quando trabalhamos com uma base de dados real, precisamos levar em conta o fato de podermos ter dados faltantes ou corrompidos, e, portanto, precisamos preparar a base de dados para nossa análise.</p> <p>As tarefas a serem realizadas são:</p> <p>a) Fazer a preparação dos dados para serem utilizados na análise, considerando que serão utilizadas apenas as variáveis survived, pclass, sex, age, sibsp, parch, fare, embarked</p> <p>i) Do conjunto de dados original, você deve selecionar um subconjunto apenas com as variáveis indicadas acima e a variável passId.</p> <p>ii) Você deve atribuir um valor para os NAs na variável age. Utilize algum critério razoável, por exemplo, o valor médio.</p> <p>iii) Você deve remover as linhas onde ainda estiverem faltando dados, depois de atribuir o valor para os NAs de age. Poucas linhas estarão ainda com dados faltantes.</p> <p>b) Você deve criar um modelo onde survived será uma função das demais variáveis.</p>

		<p>c) Faça as análises do modelo, verificando a significância estatística das variáveis (e seus parâmetros ajustados), gráficos diagnósticos, etc.</p> <p>d) Atualize o modelo como consequência da análise realizada no item anterior.</p> <p>e) Faça as previsões da variável survived na base de dados de teste utilizando o modelo refinado, e prepare um arquivo CSV para submissão que contenha apenas duas colunas: passId, survived.</p> <p>f) Submeta seu arquivo e também o script R com todas as análises realizadas, até a criação do arquivo CSV.</p>
2	Prepare-se (Requisitos):	titanic_train.csv, titanic_test.csv
3	Características técnicas	Deve ser submetido um relatório detalhado dos ajustes e todas as análises realizadas (respondendo a todos os itens do problema); deve ser submetido também o script R com todas as análises realizadas até a criação do arquivo CSV submetido.