

# Упражнения по Machine Learning (классификация и регрессия)

*самостоятельное прорешивание.  
ответы сразу после упражнения*

---

## Задание 1. Классификация

1. Загрузите данные affairs (с лекции)
  2. Перекодируйте переменные 'religion', 'occupation', 'occupation\_husb'
  3. Разбейте на обучение/тест в пропорции 70/30
  4. Обучите модель логистической регрессии и сделайте предсказание на тестовой выборке
  5. Нарисуйте кривую precision\_recall
  6. Нарисуйте roc\_curve
  7. Рассчитайте метрику roc\_auc\_score
  8. Примените метод кроссвалидации (cv=5) и рассчитайте roc\_auc\_score для каждого шага
  9. \* Попробуйте повторить предыдущие шаги, используя для обучения не все переменные
-

## Решение 1

1. `data = pd.read_csv('affair_data.csv')`
  2. `X = pd.get_dummies(data, columns=['religious', 'occupation', 'occupation_husb'])`
  3. `X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=0)`
  4. `model = LogisticRegression()`  
`model.fit(X_train, y_train)`  
`predictions = model.predict_proba(X_test)`
  5. `prec, rec, thres = precision_recall_curve(y_test, pred[:, 1])`  
`plt.plot(rec, prec)`
  6. `fpr, tpr, thres = roc_curve(y_test, pred[:, 1])`  
`plt.plot(fpr, tpr)`
  7. `roc_auc_score(y_test, pred[:, 1])`
  8. `scores = cross_val_score(LogisticRegression(), X, y, cv=10, scoring='roc_auc')`
- 

## Задание 2. Регрессия

1. Загрузите данные с лекции ('cars.csv')
  2. Закодируйте все категориальные переменные
  3. Разбейте на обучение/тест в пропорции 70/30
  4. Обучите модель линейной регрессии (LinearRegression)
  5. Обучите модель DecisionTreeRegressor()
  6. Обучите модель DecisionTreeRegressor с параметром max\_depth=4
  7. Сравните качество всех 3х моделей по метрике R2
-

## Решение 2

1. `data = pd.read_csv('cars.csv', sep='t')`
  2. `X = pd.get_dummies(data)`  
`del X['Price']`
  3. `X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=0)`
  4. `model = LinearRegression()`  
`model.fit(X_train, y_train)`  
`predictions_logit = model.predict(X_test)`
  5. `model = DecisionTreeRegressor()`  
`model.fit(X_train, y_train)`  
`predictions_tree_0 = model.predict(X_test)`
  6. `model = DecisionTreeRegressor(max_depth=4)`  
`model.fit(X_train, y_train)`  
`predictions_tree_1 = model.predict(X_test)`
  7. Рассчитайте `r2_score` для всех полученных ранее предсказаний и `y_test`
-