

Hoja informativa: Detección de anomalías

Práctica

```
# Obtener la lista de valores atípicos de la columna
boxplot = plt.boxplot(df['column'].values)
outliers = list(boxplot["fliers"][0].get_data()[1])
```

```
# Entrenar el bosque de aislamiento para la detección de anomalías
from sklearn.ensemble import IsolationForest
isolation_forest = IsolationForest(n_estimators=100)
isolation_forest.fit(data)

# Obtener la estimación de una anomalía.
# Valores de -0.5 a 0.5. Una estimación más baja indica una mayor probabilidad de que la o
bservación sea un valor atípico.
anomaly_scores = isolation_forest.decision_function(data)

# Obtener la predicción de una anomalía. -1: anomalía, 1: observación normal
estimator = isolation_forest.predict(data)

# Entrenar y obtener predicciones
estimator = isolation_forest.fit_predict(data)
```

```
# Método de detección de anomalías basado en KNN

from pyod.models.knn import KNN
model = KNN()
model.fit(data)

# Predecir anomalías: 1: anomalía, 0: observación normal
predictions = model.predict(data)
```

Teoría

Las **anomalías**, o **outliers en inglés**, son observaciones con propiedades atípicas (es decir, que se desvían de la tendencia normal). Los valores atípicos indican un problema

en los datos o que algo está fuera de lo normal.