

ADPI, 2º MIT

PRÁCTICA: ANOMALY DETECTION

CURSO ACADÉMICO 16/17

1. Introducción

En detección de anomalías una solución al problema es estimar la función densidad de probabilidad para luego estimar un umbral por debajo del cuál una muestra o evento se considera anómalo. Así en esta práctica se pide primero estimar una función de densidad de probabilidad asumiendo que es Gaussiana. Para luego, con un conjunto de entrenamiento, determinar un umbral. El proceso se repetirá para unas muestras con más definiciones.

2. Estimación de la función densidad de probabilidad

Se pretende detectar cuándo un servidor está comportándose de forma anómala. Para ello se cuenta con dos datos, el régimen binario (Mbps) y la latencia o retraso en responder (ms). Se tienen 307 muestras de test, pares de datos de máquinas, y se sospecha que en algunos casos alguna no está funcionando correctamente mientras que la mayoría si lo está haciendo.

Ejecute

```
close all
clear all
load ex8data1
```

para cargar los datos. Si teclea

```
who
```

aparecerán las variables

```
Your variables are:
X      Xval      yval
```

Se pide

- Utilizar los datos en X para modelar la función de densidad de probabilidad (pdf)
- Escriba una función que estime la probabilidad de una multidimensional Gaussiana

```
function p = multivariateGaussian(X, mu, Sigma2)
%MULTIVARIATEGAUSSIAN Computes the probability density function of the
%multivariate gaussian distribution.
%   p = MULTIVARIATEGAUSSIAN(X, mu, Sigma2) Computes the probability
%   density function of the examples X under the multivariate gaussian
%   distribution with parameters mu and Sigma2. If Sigma2 is a matrix, it is
%   treated as the covariance matrix. If Sigma2 is a vector, it is treated
%   as the \sigma^2 values of the variances in each dimension (a diagonal
%   covariance matrix)
%
% ESCRIBIR AQUI LA FUNCION
```

Nota: le puede ser de utilidad la función bsxfun.m

- Visualice la pdf utilizando contour.m. La siguiente función puede serle de utilidad.

```
function visualizeFit(X, mu, sigma2)
%VISUALIZEFIT Visualize the dataset and its estimated distribution.
%   VISUALIZEFIT(X, p, mu, sigma2) This visualization shows you the
%   probability density function of the Gaussian distribution. Each example
%   has a location (x1, x2) that depends on its feature values.
%
[X1,X2] = meshgrid(0:.5:35);
Z = multivariateGaussian([X1(:) X2(:)],mu,sigma2);
Z = reshape(Z,size(X1));

plot(X(:, 1), X(:, 2),'bx');
```

```

hold on;
% Do not plot if there are infinities
if (sum(isinf(Z)) == 0)
    contour(X1, X2, Z, 10.^(-20:3:0)');
end
hold off;
end

```

3. Detección de Anomalías

3.1. Detección de anomalías en el caso bidimensional

Para el caso del apartado anterior, y dado un umbral, por ejemplo 10^{-4} , calcular el número de anomalías.

3.2. Cálculo de umbral en el caso bidimensional

Calcular un umbral ϵ para el que se obtenga un buen valor de $F_1 = 2PR/(P + R)$ donde P es la precisión y R es el recuerdo o *recall*. La función que se incluye en WebCT le puede ser de utilidad, donde tiene que completar el código interno. El valor de ϵ debería estar en torno a $8.99e - 05$.

3.3. Detección de anomalías en un escenario con mayor dimensión

Repita los pasos anteriores para un caso más realista y con más dimensiones, 11 en este caso. El valor de ϵ debería estar en torno a $1.38e - 18$ y se deberían de encontrar 117 anomalías.