

Benemérita Universidad Autónoma de Puebla  
Facultad de Ciencias de la Computación

# Máquinas de aprendizaje

## Reporte: Gradient boosting algorithm



# BUAP

**Docente: Abraham Sánchez López**

**Alumno**

Taisen Romero Bañuelos

**Matrícula**

202055209

## Gradient boosting algorithm

Dado que en el PDF se usó una partición de 90:10 para el modelo GBM decidí usar esta misma partición de datos para los otros dos modelos con los que se hará la comparación.

Supongo que la partición de datos de 90:10 se debe a que sólo se cuenta con 150 observaciones.

```
> str(data)
'data.frame': 150 obs. of 5 variables:
 $ Sepal.Length: num 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ Sepal.Width : num 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ Petal.Length: num 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ Petal.Width : num 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
 $ Species : num 0 0 0 0 0 0 0 0 0 0 ...
```

Para los casos en que haya pocos datos se recomienda usar esta partición pero observando los valores de Kappa obtenidos hago la hipótesis de que a veces los modelos robustos como estos (que relucen con conjuntos de datos grandes) pueden comportarse de forma no óptima, ya que como vemos, hubo algunos resultados de GBM con Kappa=1, al igual que Random Forest, que también obtuvo un Kappa=1.

```
> print(GBMModel$resample) #Kappa de Gradient Boosting
  Accuracy      Kappa Resample
1 0.9615385 0.9420935 Fold3.Rep2
2 1.0000000 1.0000000 Fold1.Rep2
3 0.9629630 0.9439834 Fold5.Rep1
4 0.9615385 0.9423503 Fold4.Rep1
5 0.9259259 0.8888889 Fold3.Rep1
6 1.0000000 1.0000000 Fold4.Rep2
7 0.8928571 0.8400000 Fold2.Rep1
8 1.0000000 1.0000000 Fold1.Rep1
9 0.9642857 0.9462572 Fold2.Rep2
10 0.9285714 0.8927203 Fold5.Rep2
> print(kappa(conf_matrix)) #Kappa de XGBoost
  Estimate Std.Err 2.5% 97.5% P-value
kappa 0.8958 0.101 0.6978 1.094 7.57e-19
> kappa_RF #Kappa de Random Forest
  Estimate Std.Err 2.5% 97.5% P-value
kappa 1 0 1 1 0
> |
```

Es posible que en el caso de Random Forest haya que ajustar un poco los parámetros o incluso los datos, aún así, podemos suponer que hubo sobreajuste. Por otro lado, XGBoos obtuvo un Kappa aparentemente peor (0.895), sin embargo, considero que es un Kappa más confiable ya que la mayoría de los remuestreos de GBM tienden al sobreajuste, por lo que podemos reforzar la idea de que XGBoost es una herramienta versátil que si bien es complicado de configurar (al menos más que los otros modelos), vale la pena por su confiabilidad.