

Métodos de Analítica II

Ensamblés

Juan Eduardo Coba Puerto

Pontificia Universidad Javeriana

Section 1

- 1 Contenido
- 2 Bias-Variance
- 3 Idea del Ensamble
- 4 Bagging: Bootstrap Aggregations
- 5 CART
- 6 Random Forests
- 7 Boosting
- 8 Gradient Boosting
- 9 Gradient Boosting
- 10 Referencias

Section 2

- 1 Contenido
- 2 Bias-Variance**
- 3 Idea del Ensamble
- 4 Bagging: Bootstrap Aggregations
- 5 CART
- 6 Random Forests
- 7 Boosting
- 8 Gradient Boosting
- 9 Gradient Boosting
- 10 Referencias

Descomposición

El error de predicción de un modelo $\hat{f}(x)$ puede descomponerse en tres elementos,

$$error = Bias^2 + Var(\hat{f}(x)) + \sigma_{\epsilon}^2,$$

donde σ_{ϵ}^2 es el error irreducible, Bias corresponde a qué tanto se aleja el modelo de la realidad y $Var(\hat{f}(x))$ qué tanto oscila el modelo al rededor de la media.

Bias-Variance

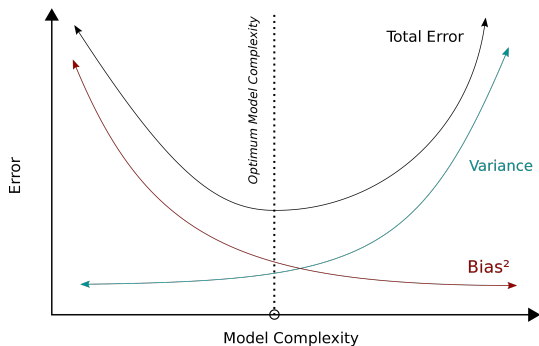


Figure: Bias-Variance Tradeoff

Bias-Variance

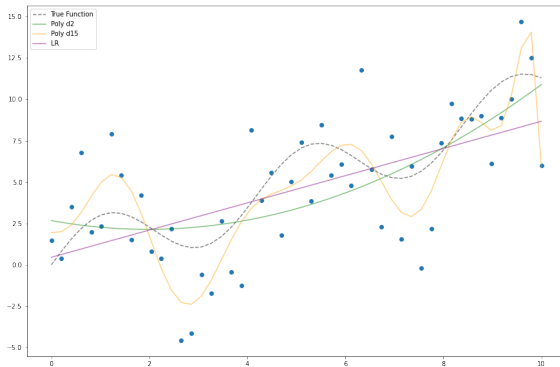


Figure: Ejemplos de Modelos con diferentes niveles de sesgo y varianza

Bias-Variance Viz!

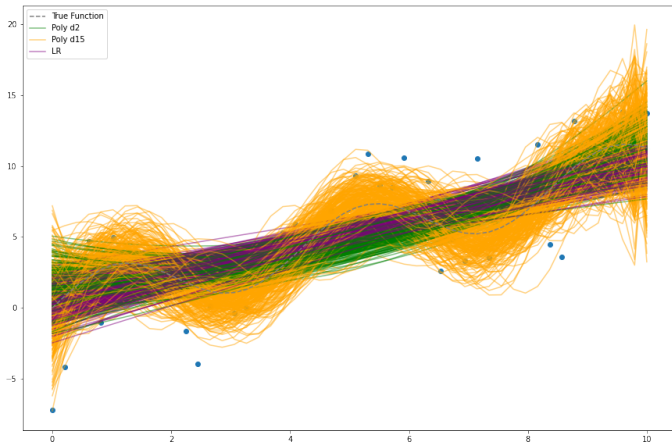


Figure: Lo que realmente significa la varianza...

Section 3

- 1 Contenido
- 2 Bias-Variance
- 3 Idea del Ensemble**
- 4 Bagging: Bootstrap Aggregations
- 5 CART
- 6 Random Forests
- 7 Boosting
- 8 Gradient Boosting
- 9 Gradient Boosting
- 10 Referencias

Diferentes modelos de clasificación

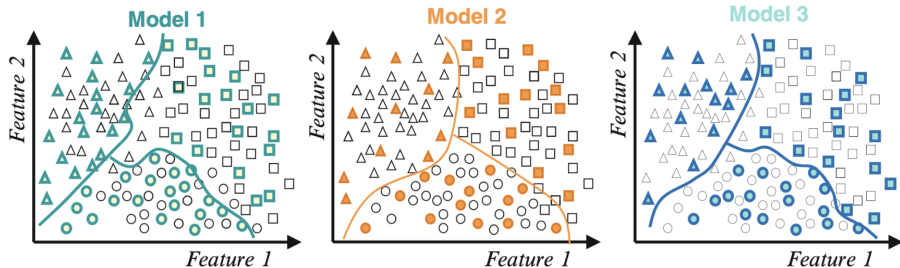


Figure: Tres fronteras de decisión diferentes

(Zhang and Ma, 2012)

Diferentes modelos de clasificación

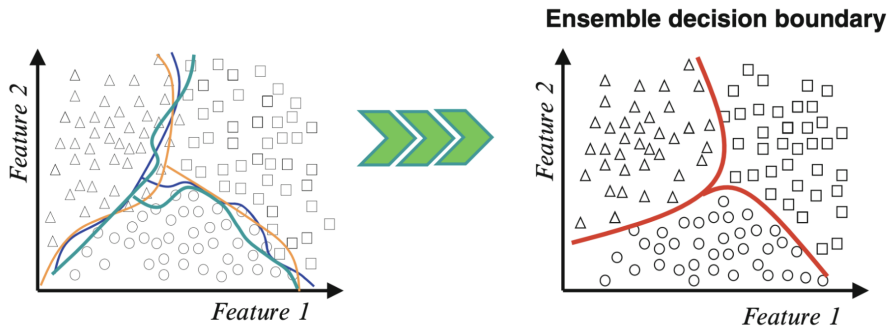


Figure: Tres fronteras de decisión diferentes... se combinan en una frontera suavizada

(Zhang and Ma, 2012)

Diferentes modelos de clasificación

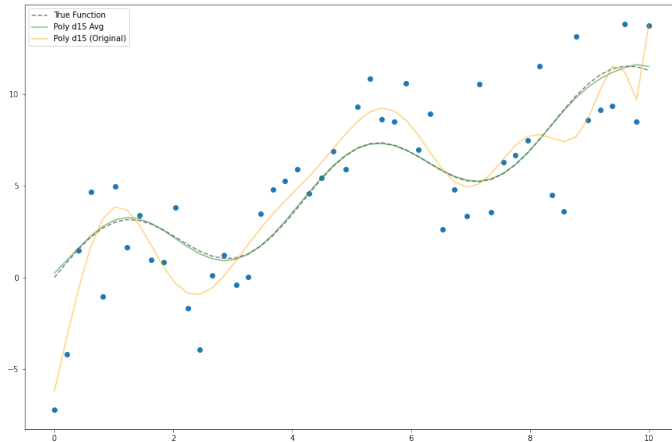


Figure: ¡Como cambió! Bajo sesgo (ya tenía) y varianza menor

Section 4

- 1 Contenido
- 2 Bias-Variance
- 3 Idea del Ensamble
- 4 Bagging: Bootstrap Aggregations**
- 5 CART
- 6 Random Forests
- 7 Boosting
- 8 Gradient Boosting
- 9 Gradient Boosting
- 10 Referencias

Bagging

Entrenar modelos en diferentes submuestras del conjunto de entrenamiento.

Dada una base de entrenamiento \mathbf{Z} conformada por parejas como $(x_i, y_i)_{i=1, \dots, N}$,

$$\mathbf{Z} = \left\{ (x_1, y_1), (x_2, y_2), \dots, (x_N, y_N) \right\}$$

Normalmente se entrenaría un modelo,

$$\hat{f}_{\mathbf{Z}}(x)$$

Bootstrap Sampling

1. Podría entrenarse el mismo modelo y obtener "formas" diferentes utilizando *muestras bootstrap*.
2. Generaría diversidad en el desempeño y particularmente en los errores de predicción.

Bootstrap Sampling

Tomar una muestra con reemplazamiento de la base de datos original,

$$\mathbf{Z}_b = \left\{ (x_{1b}, y_{1b}), (x_{2b}, y_{2b}), \dots, (x_{Nb}, y_{Nb}) \right\},$$

entrenar un modelo en esa muestra,

$$\hat{f}_b.$$

Se repite este proceso B veces.

Bootstrap Sampling



Figure: Ejemplo de Bootstrap Sampling
(Galdi and Tagliaferri, 2018)

Bagged Predictor

Como se tiene un conjunto de modelos entrenados en cada submuestra,

$$\left\{ \hat{f}_b \right\}_{b=1,2,\dots,B},$$

Entonces pueden combinarse las predicciones como,

$$\hat{f}_{\text{bagging}}(x) = \frac{1}{B} \sum_{b=1}^B \hat{f}_b(x)$$

Bootstrap Sampling

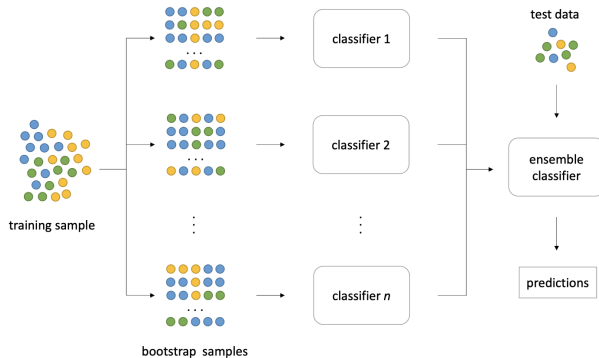


Figure: Training $n(B)$ models on bootstrap samples.

(Galdi and Tagliaferri, 2018)

Ejemplo

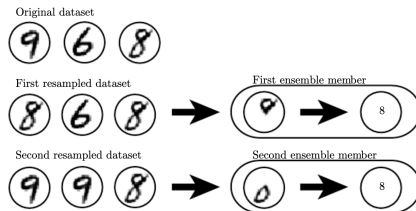


Figure: Detector de 8s. El dataset apenas contiene 6, 8 y 9. Reglas sobreajustadas.

(Goodfellow et al., 2016)

¿Dónde se usa?

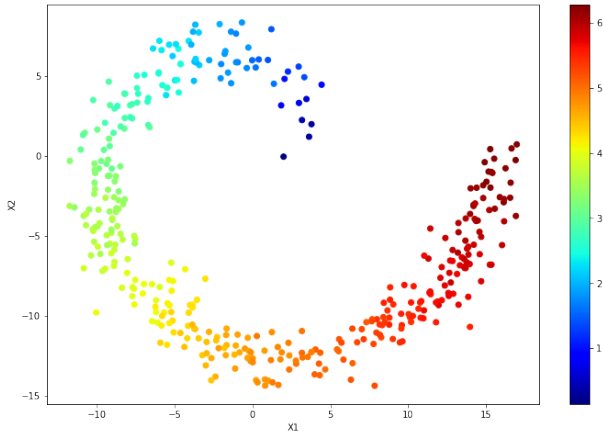
Bagging puede ser utilizado en cualquier modelo que se desee entrenar.

1. Se puede aprovechar más en modelos no lineales,
2. que detecten más particularidades de los datos (varianza).

Section 5

- 1 Contenido
- 2 Bias-Variance
- 3 Idea del Ensamble
- 4 Bagging: Bootstrap Aggregations
- 5 CART**
- 6 Random Forests
- 7 Boosting
- 8 Gradient Boosting
- 9 Gradient Boosting
- 10 Referencias

Repaso de CART



Interpretación

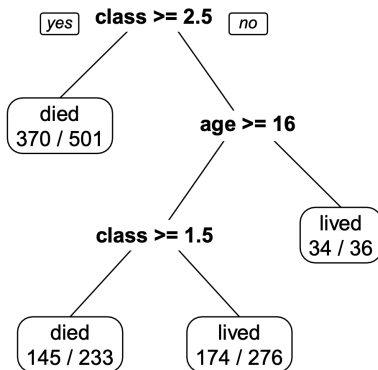


Figure: Interpretación CART Titanic
(Varian, 2014)

”Precauciones”

Tocaba tener cuidado con...

1. No dejar crecer mucho el árbol (overfitting).
2. No dejar crecer lo suficiente el árbol (underfitting).

¿Cómo sabíamos hasta dónde dejar crecer el árbol?

Section 6

- 1 Contenido
- 2 Bias-Variance
- 3 Idea del Ensamble
- 4 Bagging: Bootstrap Aggregations
- 5 CART
- 6 Random Forests**
- 7 Boosting
- 8 Gradient Boosting
- 9 Gradient Boosting
- 10 Referencias

Random Forests

Busca explotar la relación entre sesgo y varianza, al agregar *weak learners* (con mucha varianza y bajo sesgo) para conseguir una reducción en la varianza.

$$\rho\sigma^2 + \frac{1-\rho}{B}\sigma^2$$

Random Forests

Busca explotar la relación entre sesgo y varianza, al agregar *weak learners* (con mucha varianza y bajo sesgo) para conseguir una reducción en la varianza.

$$\rho\sigma^2 + \frac{1-\rho}{B}\sigma^2$$

¿Qué podríamos hacer para disminuir ρ al máximo?

Random Forest Algorithm

Algorithm Random Forest

for $b \leftarrow 1$ **to** B **do**

1. $Z_b \leftarrow$ Obtenga una muestra bootstrap de Z ;
2. $T_b \leftarrow$ entrene un árbol de decisión, con una selección aleatoria de $m \leq k$ features.;
3. Guarde el modelo.

end

return $\left\{ T_b(x) \right\}_{b=1, \dots, B}$;

Random Forest Algorithm

Para generar las predicciones utilizando el ensamble,

$$\left\{ T_b(x) \right\}_{b=1, \dots, B}$$

Regresión:

$$\hat{f}_{rf}(x) = \frac{1}{B} \sum_{b=1}^B T_b(x)$$

Clasificación: Si \hat{C}_b corresponde a la predicción que hace $T_b(x)$, entonces,

$$\hat{C}_{rf} = \text{votomayoritario} \left\{ \hat{C}_b \right\}_{b=1, \dots, B}$$

Section 7

- 1 Contenido
- 2 Bias-Variance
- 3 Idea del Ensamble
- 4 Bagging: Bootstrap Aggregations
- 5 CART
- 6 Random Forests
- 7 Boosting**
- 8 Gradient Boosting
- 9 Gradient Boosting
- 10 Referencias

Boosting

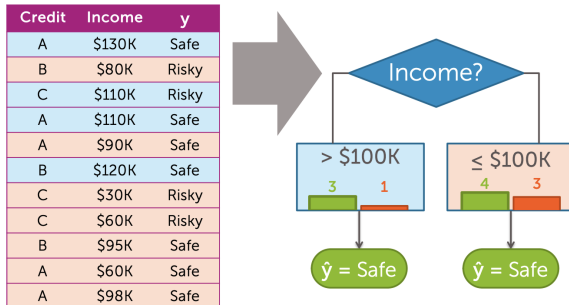
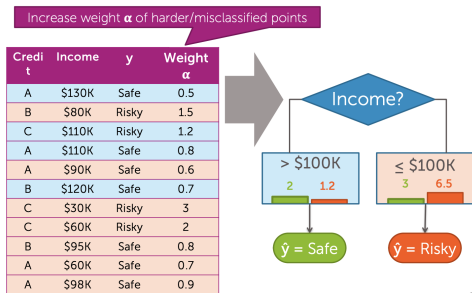


Figure: Si hacemos una predicción de riesgo (modelo sencillo)

(Ng, 2012)

Boosting



©2022 Carlos Guestrin

CS229: Machine Learning

Figure: Podríamos hacer un modelo que mejore la predicción en donde más nos equivocamos.

(Ng, 2012)

Boosting

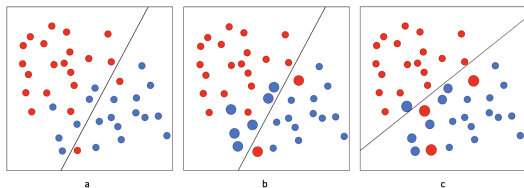


Figure: Ajustamos los pesos de las observaciones según nos equivocamos
(Galdi and Tagliaferri, 2018)

Algorithm 10.1 *AdaBoost.M1*.

1. Initialize the observation weights $w_i = 1/N$, $i = 1, 2, \dots, N$.
 2. For $m = 1$ to M :
 - (a) Fit a classifier $G_m(x)$ to the training data using weights w_i .
 - (b) Compute

$$\text{err}_m = \frac{\sum_{i=1}^N w_i I(y_i \neq G_m(x_i))}{\sum_{i=1}^N w_i}.$$
 - (c) Compute $\alpha_m = \log((1 - \text{err}_m)/\text{err}_m)$.
 - (d) Set $w_i \leftarrow w_i \cdot \exp[\alpha_m \cdot I(y_i \neq G_m(x_i))]$, $i = 1, 2, \dots, N$.
 3. Output $G(x) = \text{sign} \left[\sum_{m=1}^M \alpha_m G_m(x) \right]$.
-

Figure: Algoritmo
(Hastie et al., 2009)

$t=1$: Just learn a classifier on original data

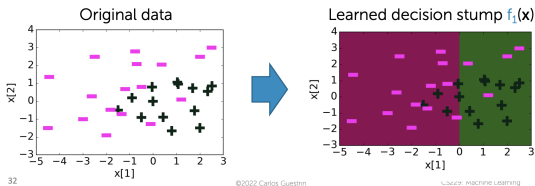


Figure: Algoritmo

(Ng, 2012)

Updating weights α_i

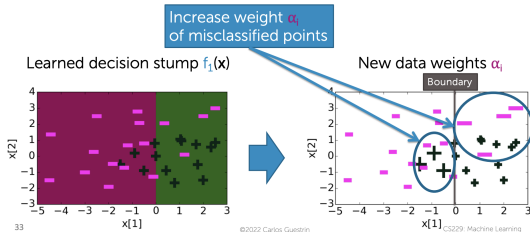
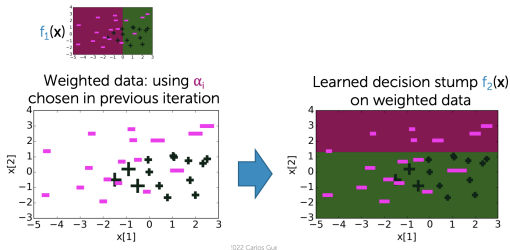


Figure: Algoritmo

(Ng, 2012)

$t=2$: Learn classifier on weighted data



©2022 Carlos Gue

Figure: Algoritmo
(Ng, 2012)

Ensemble becomes weighted sum of learned classifiers

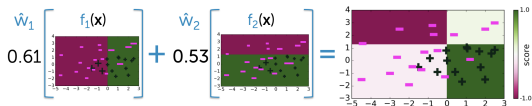


Figure: Algoritmo

(Ng, 2012)

Boosting

Este algoritmo acaba entrenando muchos modelos, de tal forma que siempre corrija los errores de los anteriores.

1. En entrenamiento con suficientes iteraciones acaba con error de 0.
2. Debe tenerse cuidado con el número de árboles para evitar overfitting (validación del modelo).

Section 8

- 1 Contenido
- 2 Bias-Variance
- 3 Idea del Ensamble
- 4 Bagging: Bootstrap Aggregations
- 5 CART
- 6 Random Forests
- 7 Boosting
- 8 Gradient Boosting**
- 9 Gradient Boosting
- 10 Referencias

Gradient Boosting

El Gradient Boosting es una mejora del algoritmo de Boosting (AdaBoost vs. XGBoost).

Section 9

- 1 Contenido
- 2 Bias-Variance
- 3 Idea del Ensamble
- 4 Bagging: Bootstrap Aggregations
- 5 CART
- 6 Random Forests
- 7 Boosting
- 8 Gradient Boosting
- 9 Gradient Boosting**
- 10 Referencias

Gradient Boosting

El Gradient Boosting es una mejora del algoritmo de Boosting (AdaBoost vs. XGBoost).

1. Se relajan las restricciones de los weak learners (de bumps a modelos cortos)
2. En lugar de predecir donde hubo error, se predicen los errores del modelo anterior.

Section 10

- 1 Contenido
- 2 Bias-Variance
- 3 Idea del Ensamble
- 4 Bagging: Bootstrap Aggregations
- 5 CART
- 6 Random Forests
- 7 Boosting
- 8 Gradient Boosting
- 9 Gradient Boosting
- 10 Referencias**

References I

- Galdi, P. and Tagliaferri, R. (2018). Data mining: accuracy and error measures for classification and prediction. *Encyclopedia of Bioinformatics and Computational Biology*, 1:431–436.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.
- Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.
- Ng, A. (2012). Cs229 lecture notes - supervised learning.
- Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2):3–28.
- Zhang, C. and Ma, Y. (2012). *Ensemble machine learning: methods and applications*. Springer.