

Sobre ajuste

Sobre ajuste es entrenar un modelo que se ajusta muy bien a los datos de entrenamiento, pero no generaliza correctamente.

Sobre ajuste

Sobre ajuste es entrenar un modelo que se ajusta muy bien a los datos de entrenamiento, pero no generaliza correctamente.

Generalizar correctamente = hacer predicciones precisas para nuevas muestras.

Sobre ajuste

Sobre ajuste es entrenar un modelo que se ajusta muy bien a los datos de entrenamiento, pero no generaliza correctamente.

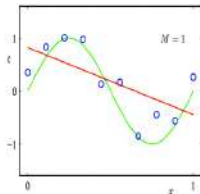
Generalizar correctamente = hacer predicciones precisas para nuevas muestras.

Causas:

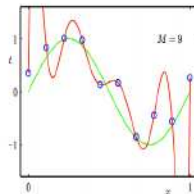
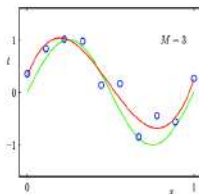
- Muestras de entrenamiento muy ruidosas
- Un modelo demasiado complejo que se ajuste al ruido
- Un conjunto de entrenamiento demasiado pequeño
- Un espacio de características de muy alta dimensión

Sobre ajuste

Regression:

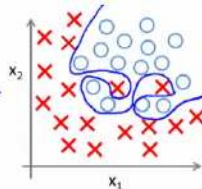
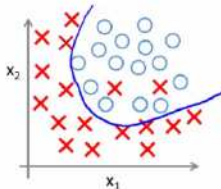
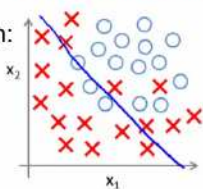


predictor too inflexible:
cannot capture pattern



predictor too flexible:
fits noise in the data

Classification:



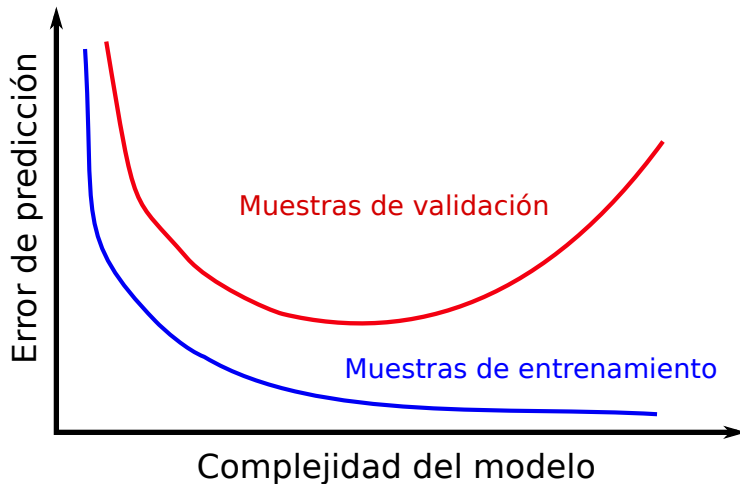
Formas de evitar el sobre ajuste

El problema del sobre ajuste requiere ser abordado desde diferentes frentes. El primero de ellos es intervenir el entrenamiento del modelo el cuál tiene varias aproximaciones:

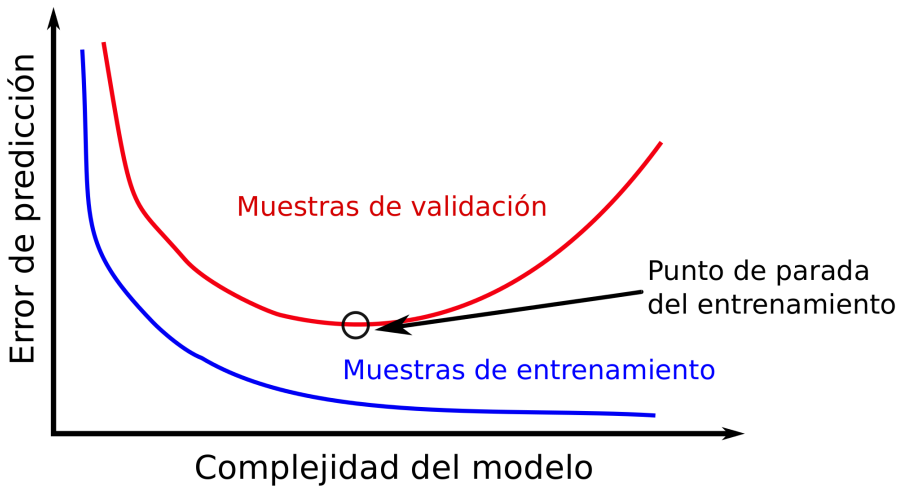
- Usar metodologías de validación apropiadas (validación cruzada, bootstrapping).
- Utilizando funciones objetivo regularizadas
- Aplicando el principio de minimización del riesgo estructural

3 de las 4 causas listada en la diapositiva anterior hacen parte de un problema conocido como la maldición de la dimensionalidad.

Metodologías de validación



Metodologías de validación



Metodologías de validación

Es necesario entonces dividir el conjunto de muestras en tres subconjuntos disjuntos: entrenamiento, validación y test:

- Entrenamiento: Es el subconjunto utilizado para ajustar los parámetros del modelo elegido.
- Validación: Es un subconjunto utilizado para evitar el sobreajuste y decidir cuándo detener el entrenamiento o seleccionar el mejor valor de algún parametro que debe ajustarse apriori (Ej: número de vecinos, número de gaussianas, tamaño del kernel, etc.).
- Test: Es un subconjunto utilizado para evaluar la capacidad de generalización del modelos entrenado.

Se deben hacer varias repeticiones usando metodologías como validación cruzada o bootstrapping, para obtener desempeño \pm IC.

Regularización

La regularización es un método matemático que permite reducir los efectos del ruido en problemas inversos imponiendo algún tipo de restricción en la función objetivo. Particularmente en el caso de la regresión lineal se determinó que cuando se presenta sobreajuste los coeficientes del polinomio tienden a aumentar en magnitud. La función objetivo regularizada está dada por:

$$E(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^N (y_i - f(\mathbf{x}_i, \mathbf{w}))^2 + \frac{\lambda}{2} \|\mathbf{w}\|^2$$

Entre mayor sea λ , mayor será la penalización por complejidad del modelo.

Regularización

dado un conjunto de datos de entrenamiento $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$, siendo \mathbf{x}_i un vector que contiene todas las características de la muestra i y y_i el correspondiente valor de la variable que se desea predecir a partir de \mathbf{x}_i , la regla de actualización de pesos regularizada para el algoritmo de gradiente estará dada por:

$$w_j = w_j - \eta \left(2\lambda w_j + 2 \sum_{i=1}^N (f(\mathbf{x}_i, \mathbf{w}) - y_i) x_{ij} \right)$$

La solución analítica regularizada para el problema de regresión está dada por:

$$\mathbf{w} = \left(\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I} \right)^{-1} \mathbf{X}^T \mathbf{Y}$$

La maldición de la dimensionalidad

La maldición de la dimensionalidad es un problema muy conocido en el campo de la minería de datos y el aprendizaje de máquina que puede ser resumido como tener un conjunto de muestras muy pequeño y un número de características muy grande para hacer algún tipo de inferencia o entrenar un modelo de aprendizaje.

La maldición de la dimensionalidad

La maldición de la dimensionalidad es un problema muy conocido en el campo de la minería de datos y el aprendizaje de máquina que puede ser resumido como tener un conjunto de muestras muy pequeño y un número de características muy grande para hacer algún tipo de inferencia o entrenar un modelo de aprendizaje.

¿Cómo atacar la maldición de la dimensionalidad?

La maldición de la dimensionalidad

La maldición de la dimensionalidad es un problema muy conocido en el campo de la minería de datos y el aprendizaje de máquina que puede ser resumido como tener un conjunto de muestras muy pequeño y un número de características muy grande para hacer algún tipo de inferencia o entrenar un modelo de aprendizaje.

¿Cómo atacar la maldición de la dimensionalidad?

Para responder la pregunta anterior debemos primero responder la siguiente ¿Cómo deseamos que sean las características?

La maldición de la dimensionalidad

La maldición de la dimensionalidad es un problema muy conocido en el campo de la minería de datos y el aprendizaje de máquina que puede ser resumido como tener un conjunto de muestras muy pequeño y un número de características muy grande para hacer algún tipo de inferencia o entrenar un modelo de aprendizaje.

¿Cómo atacar la maldición de la dimensionalidad?

Para responder la pregunta anterior debemos primero responder la siguiente ¿Cómo deseamos que sean las características?

- Discriminativas (Problema de clasificación)
- Correlacionadas con la variable a predecir (Problema de regresión)
- Independientes entre si (estadísticamente hablando)

Formas de atacar la maldición de la dimensionalidad

Teniendo en cuenta las propiedades deseadas de las características, el conducto regular que debería aplicarse sería (algunas aplican a problemas de clasificación, otras a problemas de regresión y algunas a ambos):

- Llevar a cabo análisis de varianza (ANOVA)
- Evaluar la correlación entre variables y eliminar una de las variables si se encuentra alta correlación.
- Evaluar la capacidad discriminante de las características y eliminar características de baja discriminación (C).
- Evaluar la correlación de las características con respecto a la variable a predecir y eliminar las de baja correlación (R).
- Aplicar técnicas de Selección y/o Extracción de características.

Coeficiente de correlación

El **coeficiente de correlación de Pearson** es un índice que mide la relación lineal entre dos variables aleatorias cuantitativas. Difiere de la covarianza en que es independiente de la escala de medida de las variables. Está dado por:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}$$

- Si $\rho = 1$ existe correlación positiva perfecta
- Si $0 < \rho < 1$ existe correlación positiva
- Si $\rho = 0$ no existe relación lineal
- Si $-1 < \rho < 0$ existe correlación negativa
- Si $\rho = -1$ existe correlación negativa perfecta

Fisher's Discriminant ratio

El **cociente discriminante de Fisher** es un índice que permite medir la capacidad discriminante de una o varias características bajo la suposición de que una buena característica debe proporcionar baja dispersión intra-clase y alta dispersión entre-clases.

$$F = \sum_i^C \sum_{j \neq i}^C \frac{(\mu_i - \mu_j)^2}{\sigma_i^2 + \sigma_j^2}$$

No tiene un rango de valores limitado, para hacer una evaluación objetiva sobre un conjunto de características se puede calcular el cociente de Fisher para cada una y normalizar todos los valores con respecto al máximo. Las características de mayor capacidad discriminante serán las que tengan índices por encima de 0.5.

Selección/Extracción de Características

Los métodos vistos hasta ahora evalúan las características de manera individual, sin embargo, es posible que un par de características individualmente no tenga un buen comportamiento, pero la combinación entre ellas si lo tenga. Los métodos de extracción y selección de características consideran el conjunto completo de características para hacer la evaluación.

- Selección: Utiliza una función de medición del desempeño y evalúa diferentes combinaciones de características para seleccionar el mejor subconjunto. (Fuerza bruta, búsqueda secuencial ascendente, búsqueda secuencial descendente y búsqueda secuencial flotante)
- Extracción: Su objetivo es producir un nuevo subconjunto de características que contengan la mayor parte de información de las características originales, a través de combinaciones lineales o no lineales de éstas (PCA, LDA).

Referencias

- [1] Velten K., *Mathematical Modeling and Simulation*, WILEY-VCH, 2009.
- [2] Murphy K.P., *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.
- [3] Duda R.O., Hart P.E., Stork D.G., *Pattern Classification*. 2ed, WILEY-INTERSCIENCE, 2001.
- [4] Bishop, C.M. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [5] Webb, A.R. *Statistical Pattern Recognition*, 2nd Revised edition, John Wiley & Sons Ltd, 2002.