Plan de trabajo propuesto para trabajo final de grado.

Licenciatura en Matemática. Departamento de Matemática. FaEA-UNCo

Título: Fundamentos Matemáticos del Aprendizaje Automático.

Estudiante: Nicolás Silva Nash.

Director: Dr. Luis Nowak.

Codirectora: Dra. Alejandra Perini.

Introducción

Durante las últimas décadas, el desarrollo de métodos automáticos capaces de aprender a partir de datos ha transformado profundamente tanto la investigación científica como diversas aplicaciones tecnológicas, industriales y sociales. Esta revolución ha sido impulsada por la disponibilidad masiva de datos, el aumento de la capacidad computacional y el diseño de algoritmos cada vez más potentes. Sin embargo, más allá de sus aplicaciones visibles, el aprendizaje automático descansa sobre fundamentos matemáticos profundos, cuyo estudio riguroso es esencial para comprender las verdaderas capacidades y limitaciones de los algoritmos.

La Teoría del Aprendizaje Estadístico constituye un ejemplo paradigmático de integración de ramas fundamentales de la matemática —probabilidad, teoría de la medida, estadística, análisis de la convergencia y optimización— en el estudio de un problema aplicado: el aprendizaje a partir de datos. Esta teoría no solo permite evaluar el comportamiento de los algoritmos existentes, sino que también guía el diseño de nuevos métodos con mejores propiedades de generalización.

El aprendizaje a partir de datos plantea una de las preguntas centrales de la ciencia contemporánea: ¿es posible inferir reglas generales, confiables y estables, a partir de observaciones finitas y sujetas al azar? Esta cuestión, que atraviesa campos tan diversos como la estadística, la inteligencia artificial y la teoría de la información, ha dado lugar a un cuerpo teórico profundo y estructurado: la *Teoría del Aprendizaje Estadístico* (Statistical Learning Theory, SLT).

Este trabajo se enmarca en una perspectiva matemática rigurosa, con el propósito de construir una comprensión sólida del aprendizaje como fenómeno teórico. Se abordarán los

principios que permiten garantizar que el aprendizaje desde datos no solo sea posible, sino también controlable y explicable desde los fundamentos de la matemática. Específicamente, se propone explorar los fundamentos matemáticos del aprendizaje desde datos a través de lo que aquí denominamos su **tríada conceptual**: *riesgo*, *complejidad* y *convergencia*. Estos tres ejes condensan los desafíos centrales que enfrenta cualquier algoritmo de aprendizaje supervisado: medir y minimizar el error, controlar la capacidad de generalización, y garantizar un comportamiento estable al aumentar el tamaño de la muestra.

La noción de **riesgo** formaliza el desempeño de una función predictiva, entendida como su capacidad de acertar en nuevas observaciones provenientes de una distribución desconocida. Dado que esta distribución no es accesible, se introduce el riesgo empírico, cuya relación con el riesgo verdadero constituye uno de los problemas más estudiados de la SLT.

La **complejidad**, en tanto, remite a la capacidad expresiva de los modelos que se utilizan para aprender: clases de funciones más complejas pueden ajustar mejor los datos, pero también son más propensas a sobreajustar y generalizar peor. Medidas como la dimensión de Vapnik–Chervonenkis (VC) permiten cuantificar esta capacidad de forma rigurosa.

Finalmente, la **convergencia** aparece como el principio unificador de la teoría: se trata de estudiar bajo qué condiciones, y con qué velocidad, los algoritmos de aprendizaje logran aproximarse al clasificador óptimo a medida que reciben más datos. Resultados como las desigualdades de concentración, los principios de convergencia uniforme y los teoremas de consistencia estadística constituyen herramientas centrales para este análisis.

1 Marco Teórico: Fundamentos del aprendizaje supervisado

En el aprendizaje supervisado, el objetivo es aprender una función que relacione entradas $X \in \mathcal{X}$ (espacio de características) con salidas $Y \in \mathcal{Y}$ (etiquetas o clases). Se asume que los datos disponibles provienen de una distribución de probabilidad conjunta desconocida $P(X, Y)$, y que se cuenta con una muestra finita:

$$(X_1, Y_1), \ldots, (X_n, Y_n) \sim P \quad \text{i.i.d.}$$

El objetivo es construir una función $f : \mathcal{X} \to \mathcal{Y}$ tal que, al aplicarla a nuevas observaciones X, produzca predicciones $f(X)$ lo más cercanas posible a los verdaderos valores Y. Para evaluar el desempeño de f, se introduce una función de pérdida $\ell : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_+$, que mide el costo de predecir $f(X)$ cuando el valor verdadero es Y.

El *riesgo esperado* (o verdadero) se define como:

$$R(f) = \mathbb{E}_{(X,Y)\sim P}[\ell(f(X), Y)]$$

Como la distribución P es desconocida, no es posible calcular $R(f)$ directamente. En su lugar, se utiliza una estimación basada en los datos disponibles, conocida como *riesgo empírico*:

$$R_n(f) = \frac{1}{n} \sum_{i=1}^{n} \ell(f(X_i), Y_i)$$

El aprendizaje se convierte, entonces, en un problema de *optimización*: encontrar la función $f \in \mathcal{F}$ (dentro de una clase de funciones dada) que minimice el riesgo empírico, con la esperanza de que también minimice el riesgo verdadero. A medida que aumenta el tamaño de la muestra n, esperamos que $R_n(f) \to R(f)$, es decir, que las funciones seleccionadas por el algoritmo *converjan* hacia una solución ideal. Esta convergencia es el corazón del análisis matemático del aprendizaje.

Un algoritmo de aprendizaje *generaliza bien* si su rendimiento sobre datos nuevos (no vistos) es similar al observado durante el entrenamiento. Matemáticamente, esto se expresa exigiendo que la diferencia $|R(f) - R_n(f)|$ sea pequeña con alta probabilidad.

Esta propiedad se analiza mediante herramientas probabilísticas como las *desigualdades de concentración* (Hoeffding, Chernoff), y se complementa con resultados de *convergencia casi segura* y *en probabilidad*, que aseguran que el comportamiento del algoritmo mejora conforme crece la muestra.

Vapnik (1995, 1998) propuso el principio de *minimización del riesgo empírico*, que sugiere controlar la complejidad del conjunto de hipótesis para evitar el sobreajuste. Por su parte, Devroye et al. (1996) demostraron que ciertos algoritmos, como el de los *k-vecinos más cercanos*, son *universalmente consistentes*, es decir, convergen al clasificador de Bayes bajo condiciones generales.

El análisis del aprendizaje no depende únicamente de los datos, sino también de la clase de funciones $\mathcal{F}$ que se utiliza. Una clase demasiado amplia puede sobreajustar los datos, mientras que una clase muy restringida puede ser incapaz de aproximar la verdadera relación entre X y Y.

La SLT introduce medidas de *capacidad del modelo*, como la *dimensión de Vapnik–Chervonenkis* (VC), que permite cuantificar la expresividad de $\mathcal{F}$. Este concepto está en la base de muchos resultados de *convergencia uniforme*, que permiten establecer cotas del tipo:

$$\sup_{f \in \mathcal{F}} |R(f) - R_n(f)| \leq \epsilon \quad \text{con alta probabilidad}$$

La *decomposición del error total* en error de aproximación y error de estimación permite analizar con precisión los factores que influyen en el rendimiento del modelo:

$$R(f_n) - R(f^*) = \underbrace{R(f_n) - R(f_{\mathcal{F}})}_{\text{Estimación}} + \underbrace{R(f_{\mathcal{F}}) - R(f^*)}_{\text{Aproximación}}$$

Este análisis está íntimamente ligado a resultados de teoría de la convergencia, que garantizan el comportamiento asintótico de los algoritmos.

Entre los principales resultados matemáticos que serán abordados en este trabajo, se destacan:

- **Desigualdad de Hoeffding**: para variables acotadas, permite obtener cotas de concentración del riesgo empírico.
- **Principio de convergencia uniforme** (Vapnik, 1995): garantiza que la convergencia del riesgo empírico al verdadero ocurre uniformemente sobre $\mathcal{F}$, bajo condiciones sobre la capacidad.
- **Teoremas de consistencia universal**: como los demostrados por Stone (1977) para el clasificador k-NN.

- **Principio de minimización del riesgo empírico (SRM)**: propone seleccionar funciones que equilibren riesgo empírico bajo con baja complejidad.

Estos resultados se formulan y demuestran usando herramientas de probabilidad, teoría de la medida, análisis real, teoría de la optimización y estadística matemática.

2 Objetivos

2.1 Objetivo General

Analizar y formalizar los fundamentos matemáticos del aprendizaje estadístico, a partir del estudio riguroso del marco de clasificación supervisada, con énfasis en las nociones de generalización, consistencia, capacidad de hipótesis y sus consecuencias teóricas en el diseño y análisis de algoritmos de aprendizaje.

2.2 Objetivos Particulares

1. Estudiar y comprender los modelos probabilísticos que dan origen a los problemas de clasificación, con énfasis en la generación de datos a partir de distribuciones de probabilidad $P(X, Y)$ y su interpretación como procesos de muestreo i.i.d.
2. Definir y analizar el riesgo verdadero y el riesgo empírico, junto con los conceptos de consistencia del algoritmo de aprendizaje y generalización de la función aprendida.
3. Presentar y discutir los principales resultados teóricos que vinculan la capacidad del conjunto de hipótesis con el error de generalización, incluyendo:
   - El teorema de consistencia del clasificador k-NN (Stone, 1977),
   - Las desigualdades de concentración (Hoeffding, Chernoff),
   - La descomposición del error en componentes de estimación y aproximación.
4. Estudiar formalmente la dimensión VC y el principio de convergencia uniforme como herramientas para acotar el error de generalización y justificar el principio de minimización del riesgo empírico.
5. Contrastar el marco clásico de la estadística paramétrica con el enfoque agnóstico del aprendizaje estadístico, en el que no se asume conocimiento previo sobre la forma funcional de la distribución subyacente.
6. Analizar el rol del clasificador de Bayes como solución teórica óptima en clasificación supervisada, y su relación con los algoritmos prácticos que intentan aproximarlo.

3 Metodología.

El desarrollo del trabajo se abordará desde una perspectiva teórico-matemática, estructurada en etapas sucesivas que permitan construir una comprensión rigurosa y progresiva de los fundamentos del aprendizaje estadístico. La metodología estará centrada en el estudio analítico de definiciones, proposiciones, teoremas y sus respectivas demostraciones, así como en la construcción de ejemplos ilustrativos que permitan visualizar y contextualizar los resultados

teóricos. Se organizarán reuniones periódicas entre tesista y directores para trabajar en el estudio de los temas y desarrollo del trabajo final.

References

[1] C. Bishop (2006). *Pattern Recognition and Machine Learning*.

[2] Devroye, L., Györfi, L., & Lugosi, G. (1996). *A Probabilistic Theory of Pattern Recognition*. Springer-Verlag, New York.

[3] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, **58**(301), 13–30.

[4] Luxburg, U. von, & Schölkopf, B. (2008). *Statistical Learning Theory: Models, Concepts, and Results*. In *Dagstuhl Seminar Proceedings 07161*.

[5] Abu-Mostafa (2012). *Learning From Data*. Preprint

[6] Stone, C. J. (1977). Consistent nonparametric regression. *The Annals of Statistics*, **5**(4), 595–620.

[7] Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag, New York.

[8] Vapnik, V. N. (1998). *Statistical Learning Theory*. Wiley-Interscience, New York.