

# Machine Learning

---

Fernando Lozano

e-mail: flozano@uniandes.edu.co

Oficina: ML 324. Horas de Atención: Lunes 3:30-4:30, Miércoles 2:00-3:00

---

Martes 2:00 - 3:20 pm Salón: AU 202

Viernes 2:00 - 3:20 pm Salón: AU 202

---

## 1. Descripción

Machine Learning estudia la solución por computador de problemas mediante algoritmos que *aprenden* de su experiencia. Este tipo de técnicas son hoy en día muy exitosas en contextos en los que una solución programada no es posible. El objetivo de este curso es proveer al estudiante con las herramientas necesarias para la aplicación de técnicas de Machine Learning a la solución de problemas prácticos. Se estudian aspectos generales necesarios para la solución de cualquier problema como pre-procesamiento de datos, evaluación y selección de modelo. Se estudian las técnicas más populares para aprendizaje supervisado tales como Redes Neuronales y Support Vector Machines.

## 2. Contenido

1. Revisión Optimización y Probabilidad.
2. Introducción. Tipos de aprendizaje.
3. Regresión Lineal.
4. Aprendizaje supervisado. [Devroye et al., 1997] capítulo 2, [Bishop, 1995] capítulo 1.
5. Perceptrón y redes neuronales [Bishop, 1995], capítulos 3 y 4. [Bishop, 2006], capítulos 4 y 5.
6. Selección de modelo. [Kearns et al., 1997, Bartlett et al., 2000]
7. Combinación de clasificadores [Schapire and Freund, 2012]
8. Support Vector Machines y aprendizaje con kernels.
  - (a) Introducción tutorial [Burges, 1998]
  - (b) Teoría de Kernels [Scholkopf and Smola, 2002] capítulo 2.
  - (c) SVMs. [Scholkopf and Smola, 2002] capítulo 7.
  - (d) Regresión con Kernels [Scholkopf and Smola, 2002] capítulo 13.
9. Preprocesamiento y selección de características. [Bishop, 1995] capítulo 8.
10. Problemas multiclase, problemas no balanceados. [Dietterich and Bakiri, 1995, Allwein et al., 2000], [Zadrozny et al., 2003, Domingos, 1999].

11. Utilización de datos sin etiquetar <sup>\*</sup>
  - (a) Aprendizaje semisupervisado. [Chapelle et al., 2010]
  - (b) Aprendizaje activo [Dasgupta, 2004, Freund et al., 1997].
12. Aprendizaje no supervisado (clustering)<sup>\*</sup> [MacKay, 2003].
13. Aprendizaje por refuerzo <sup>\*</sup> [Sutton and Barto, 1998]

### 3. Evaluación

- Tareas 50 %
- Examen parcial 20 %
- Proyecto 30 %

### 4. Algunas reglas

- Las tareas deben realizarse de manera **INDIVIDUAL**. Esto quiere decir que aunque es válido discutir los problemas con sus compañeros, la solución debe ser de su completa autoría. En particular está prohibido resolver problemas en grupo, y por supuesto copiar literalmente la solución y/o procedimiento desarrollado por otro estudiante, u obtener soluciones del internet u otros medios <sup>\*\*</sup> Cualquier transgresión a esta regla se considerará **FRAUDE** y se reportará sin excepciones.
- Cada tarea tendrá una fecha y hora de entrega predeterminada. Usted dispone de un “presupuesto” de 5 días de retardo que puede utilizar libremente sin incurrir en ninguna penalización (por ejemplo usted puede entregar la primera tarea 3 días tarde y la segunda dos días tarde y las demás a tiempo). Días adicionales de retardo se penalizarán restando una unidad a la nota final por cada día de retardo.
- La calificación de las tareas se hará de manera selectiva de acuerdo a criterio del profesor.

## Referencias

- [Allwein et al., 2000] Allwein, E. L., Schapire, R. E., and Singer, Y. (2000). Reducing multiclass to binary: A unifying approach for margin classifiers. *Journal of machine learning research*, 1.
- [Bartlett et al., 2000] Bartlett, P., Boucheron, S., and Lugosi, G. (2000). Model selection and error estimation. In *Proceedings of the thirteenth annual conference on computational learning theory*.

---

<sup>\*</sup>Secciones opcionales, se cubrirán dependiendo del tiempo disponible.

<sup>\*\*</sup>En algunas ocasiones usted puede obtener crédito parcial por una solución obtenida de alguna fuente, siempre y cuando esté apropiadamente referenciada en el texto y usted la entienda por completo. Esta regla aplicará de acuerdo a criterio del profesor.

- [Bishop, 1995] Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford University Press.
- [Bishop, 2006] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.
- [Burges, 1998] Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2:121–167.
- [Chapelle et al., 2010] Chapelle, O., Schölkopf, B., and Zien, A. (2010). *Semi-Supervised Learning*. The MIT Press, 1st edition.
- [Dasgupta, 2004] Dasgupta, S. (2004). Analysis of a greedy active learning strategy. In *Neural information processing systems (NIPS)*.
- [Devroye et al., 1997] Devroye, L., Györfi, L., and Lugosi, G. (1997). *A Probabilistic Theory of Pattern Recognition (Stochastic Modelling and Applied Probability)*. Springer.
- [Dietterich and Bakiri, 1995] Dietterich, T. G. and Bakiri, G. (1995). Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research*, 2.
- [Domingos, 1999] Domingos, P. (1999). A general method for making classifiers cost sensitive. In *Proceedings of the 5th International Conference on Knowledge Discovery and Data Mining*.
- [Freund et al., 1997] Freund, Y., Seung, H. S., Shamir, E., and Tishby, N. (1997). Selective sampling using the query by committee algorithm. *Machine Learning*, 28(2-3):133–168.
- [Kearns et al., 1997] Kearns, M., Mansour, Y., Ng, A., and Ron, D. (1997). An experimental and theoretical comparison of model selection methods. *Machine Learning*, 27(1).
- [MacKay, 2003] MacKay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge University Press.
- [Schapire and Freund, 2012] Schapire, R. E. and Freund, Y. (2012). *Boosting: Foundations and Algorithms*. The MIT Press.
- [Schölkopf and Smola, 2002] Schölkopf, B. and Smola, A. (2002). *Learning with kernels*. MIT Press.
- [Sutton and Barto, 1998] Sutton, R. S. and Barto, A. G. (1998). *Reinforcement Learning: an introduction*. MIT press.
- [Zadrozny et al., 2003] Zadrozny, B., Langford, J., and Abe, N. (2003). Cost sensitive learning by cost-proportionate example weighting. In *Proceedings of the Third IEEE International Conference on Data Mining*.