

PROGRAMA DE CURSO

Código	Nombre		
MA5705	Optimización para el análisis de datos		
Nombre en Inglés			
Optimization for data science			
SCT	Horas de Cátedra	Horas Docencia Auxiliar	Horas de Trabajo Personal
6	4	0	6
Requisitos			Carácter del Curso
MA3711 o MA3701			Electivo de carrera, magister y doctorado
Resultados de Aprendizaje			
<p>Este curso es de carácter introductorio, presentando lo esencial sobre optimización convexa y métodos numéricos con el objetivo de entender, aplicar y adaptar las principales técnicas de aprendizaje estadístico. Asimismo, se estudiará como las técnicas de aprendizaje han permitido mejorar herramientas en optimización y en investigación de operaciones.</p> <p>El enfoque teórico del curso es basado principalmente en optimización convexa, mientras que el aspecto práctico considera programación en Python.</p> <p>Al final del curso, el alumno conocerá el estado del arte en cuanto a los métodos de optimización utilizados para entrenar algoritmos de aprendizaje de máquinas, y será capaz de aplicar estos conocimientos al utilizar algoritmos presentes en librerías como también entender o proponer nuevos métodos.</p>			

Metodología Docente	Evaluación General
<p>2 clases semanales dictadas por el profesor.</p> <p>Se entregarán tareas que permitan conectar la teoría con la práctica.</p>	<p>La evaluación será en aspectos tanto teóricos como prácticos durante el desarrollo del curso e incluirá:</p> <ul style="list-style-type: none"> • Presentaciones: • Asistencia y participación en presentaciones de otros alumnos • Participación en discusión de las técnicas y discusiones grupales en base a material de grupo de lectura • Tareas computacionales

Unidades Temáticas

Número	Nombre de la Unidad	Duración en Semanas
1	Nociones básicas del análisis convexo y dualidad	3
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
<ul style="list-style-type: none"> - Funciones convexas; sus principales propiedades - El Subdiferencial de una función convexa - Conjugada de Fenchel - Dualidad de Fenchel-Rockafellar - Dualidad Lagrangeana - Operadores proximales - Ejemplos en regresión, clasificación y análisis de datos en general 	<ul style="list-style-type: none"> • Se introducen las nociones básicas del análisis convexo en dimensión finita como funciones y conjuntos convexas, conjuntos de recesión, subdiferenciales, etc. • El alumno sabrá calcular tanto el dual convexo como el de perturbaciones (Fenchel-Rockafellar) de un problema convexo. • El alumno sabrá aplicar estos conocimientos a ejemplos concretos que se usan en análisis de datos. 	1, 2, 3 y 7
Número	Nombre de la Unidad	Duración en Semanas
2	Algoritmos de optimización de primer orden	4
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
<ul style="list-style-type: none"> • Riesgo empírico y riesgo esperado • Algoritmos de descenso • El método del gradiente y su complejidad • Métodos acelerados y de quasi-Newton • Algoritmos para modelos regularizados no-diferenciables <ul style="list-style-type: none"> ○ Algoritmo Proximal ○ ISTA y variantes 	<ul style="list-style-type: none"> • Se revisan contenidos de métodos de primer orden con énfasis en puntos relevantes para las aplicaciones, como tasas de convergencia, condicionamiento y búsqueda lineal • Se introducen variantes recientes de los métodos de primer orden básicos • El alumno entiende la importancia de la regularización de problemas de aprendizaje y conoce los algoritmos especializados para este tipo de problemas 	7, 8, 11, 12, 15

Número	Nombre de la Unidad	Duración en Semanas
3	Algoritmos estocásticos: SGD y variantes	3
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
<p>Se estudian variantes estocásticas del método del gradiente, entre las cuales:</p> <ul style="list-style-type: none"> • Coordinate descent • Stochastic Gradient Descent • Mini-batch Gradient Descent • Adagrad 	<ul style="list-style-type: none"> • El alumno comprende cuales son las motivaciones y los beneficios de introducir estocasticidad en los problemas de optimización en aprendizaje • El alumno conoce las distintas variantes de los métodos estocásticos, su deducción y los problemas que busca solucionar 	9, 16,18, 19

Así como algunas técnicas para mejorar su desempeño en problemas de Machine Learning	<ul style="list-style-type: none"> El alumno comprende y sabe identificar el comportamiento característico de cada uno de ellos, y sus principales propiedades teóricas 	
--	--	--

Número	Nombre de la Unidad	Duración en Semanas
4	Métodos de Optimización en Deep Learning	3
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
<p>Se estudian los aspectos específicos al problema de minimización que interviene en el entrenamiento de redes neuronales. Los siguientes conceptos son vistos en detalle</p> <ul style="list-style-type: none"> El método de 'Backpropagation' El problema de los 'vanishing gradients' y algunos remedios Learning rates: su importancia y estrategias de ajuste Mínimos locales y globales en redes neuronales 	<ul style="list-style-type: none"> Se repasa brevemente la formulación matemática de una red neuronal Se formula el método de <i>backpropagation</i> como un algoritmo de descenso En base a sus conocimientos de algoritmos de descenso tanto deterministas como estocásticos, el alumno es capaz de entender en profundidad el método de <i>backpropagation</i>, así como el origen de sus dificultades más frecuentes Al final de la unidad, el alumno está en condiciones de continuar autónomamente su exploración del estado del arte en el área, incluyendo las innovaciones más recientes 	12,18

Número	Nombre de la Unidad	Duración en Semanas
5	Optimización paralela y distribuida	2
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
<ul style="list-style-type: none"> Se estudian variantes paralelas y distribuidas de algoritmos presentados en unidades anteriores Se estudian en mayor detalle los algoritmos más utilizados en ambientes paralelos y distribuidos, como los algoritmos ADMM y SDCA 	<ul style="list-style-type: none"> El alumno domina conceptos básicos de optimización paralela y distribuida, como sincronismo y asincronismo El alumno conoce las variantes paralelas de los algoritmos más utilizados en ML, y es capaz de deducir paralelizaciones efectivas de algoritmos similares 	5, 6, 14

Bibliografía

- 1) Borwein, Jonathan M., and Lewis, Adrian S., *Convex Analysis and Non Linear Optimization*, Springer, NY, 2000.
- 2) Hiriart Urruty, Jean-Baptiste, and Lemarechal, Claude, *Fundamentals of Convex Analysis*, Springer Science & Business Media, 2004.
- 3) Rockafellar, R. Tyrrell, *Convex Analysis*, Princeton Mathematical Series, 1997.
- 4) Bach, Francis, et al. "Optimization with sparsity-inducing penalties." *Foundations and Trends in Machine Learning* 4.1 (2012): 1-106.
- 5) Bertsekas, Dimitri P., and John N. Tsitsiklis. *Parallel and distributed computation: numerical methods*. Vol. 23. Englewood Cliffs, NJ: Prentice hall, 1989.
- 6) Boyd, Stephen, et al. "Distributed optimization and statistical learning via the alternating direction method of multipliers." *Foundations and Trends in Machine Learning* 3.1 (2011): 1-122.
- 7) Boyd, Stephen, and Vandenberghe, Lieven, *Convex Optimization*, Cambridge Universi Press, 2004.
- 8) Combettes, Patrick L., and Valérie R. Wajs. "Signal recovery by proximal forward-backward splitting." *Multiscale Modeling & Simulation* 4.4 (2005): 1168-1200.
- 9) Fercoq, Olivier, and Peter Richtárik. "Accelerated, parallel, and proximal coordinate descent." *SIAM Journal on Optimization* 25.4 (2015): 1997-2023.
- 10) Nesterov, Yu. "Gradient methods for minimizing composite functions." *Mathematical Programming* 140.1 (2013): 125-161.
- 11) Parikh, Neal, and Stephen Boyd. "Proximal algorithms." *Foundations and Trends in Optimization* 1.3 (2014): 127-239.
- 12) Sra, Suvrit, et al., *Optimization for Machine Learning*, The MIT Press, 2012.
- 13) Takác, Martin, et al. "Mini-Batch Primal and Dual Methods for SVMs." *ICML* (3). 2013.
- 14) Takác, Martin, Jakub Marecek, and Peter Richtárik. "Distributed Coordinate Descent for Big Data Optimization."
- 15) Tibshirani, Robert. "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society. Series B (Methodological)* (1996): 267-288.
- 16) Bottou, Leon, Curtis, Frank E., and Nocedal, Jorge, "Optimization Methods for Large-Scale Machine Learning", *Siam Reviews*, 60(2):223-311, 2018.
- 17) Ruder, Sebastian. "An overview of gradient descent optimization algorithms." arXiv preprint arXiv:1609.04747 (2016).
- 18) Montavon, G., Orr, G.B., Müller, K.-R. (eds.) *NN: Tricks of the Trade*, 2nd edn. LNCS, vol. 7700, pp. 437–478. Springer, Heidelberg (2012).

Vigencia desde:	Otoño 2019
Elaborado por:	Salvador Flores, Héctor Ramírez