



Deep Learning aplicado al diseño de compuestos bioactivos
Posgrado

Carácter del curso	CURSO DE POSGRADO
Semestre en que se dicta	Segundo Hemisemestre Semestre PAR
Número de créditos	13
Carga horaria semanal (hs)	6 (teóricos) y 10 (prácticas estructuradas). Proyecto Integrador (solo para posgrado): 30 horas
Previaturas	
Cupo	30

Estructura Responsable:

Área Bioinformática - Departamento DETEMA - Facultad de Química - UdeLaR

Docentes Responsables:

Margot Paulino
Pablo García

Docentes Colaboradores:

Andrés Camilo Ballesteros
Jorge Cantero

Objetivos:

El curso se enfoca en introducir al alumno en el uso de inteligencia artificial aplicada al diseño de compuestos bioactivos. En particular se profundiza en el uso de técnicas de aprendizaje automático de la categoría de Deep Learning, introduciendo al alumno en la teoría de redes neuronales, diferentes arquitecturas de redes y su aplicación a diferentes etapas del diseño de compuestos bioactivos así como se lo introduce en las principales herramientas de aplicación de estas técnicas y entornos de ejecución de las mismas.

Fecha	V.01
		Página 1 de 6

Contenido:

Clases teóricas de 2 horas cada una

- 1 Introducción a las redes neuronales, teoría y algoritmos básicos y su aplicación a problemas de regresión y clasificación.
 - Introducción a las redes neuronales.
 - Redes neuronales biológicas.
 - Construcción de una red neuronal, tipos de capas, dimensionalidad (ancho y profundidad), visualización de la red.
 - Redes neuronales artificiales, forward propagation.
 - Algoritmos básicos :
 - Gradient Descent
 - Backpropagation
 - Vanishing Gradient
 - Funciones de activación.

- 2 Regularización y optimización (hyperparameters tuning).
 - Métodos de evaluación (Mean Squared Error, Binary Cross Entropy, Categorical Cross Entropy)
 - Hyperparámetros tuning.
 - Inicialización
 - Regularización Dropout.
 - Normalización batch.

- 3 Métodos de optimización:
 - Gradient Descent
 - Stochastic Gradient Descent
 - Momentum
 - Variable and Adaptive Learning Rates
 - Adam

- 4 Convolutional Neural Networks :
 - Concepto de convolución y kernels comúnmente utilizados.
 - Max Polling
 - Aplicación de las redes convolucionales a estructuras 3D e imágenes.

Fecha	V.01
	Página 2 de 6	

- 5 Redes neuronales deep, otras arquitecturas de uso específico:
 - RNN Recurrent Neural Networks
 - LSTM Long Short Term Memory

- 6 Introducción a los frameworks de mayor adopción en la implementación de redes neuronales deep aplicadas a diseño de compuestos bioactivos:
 - Sci-kit learn
 - Keras
 - Tensorflow

- 7 GAN Generative adversarial networks
 - Modelos generativos.
 - Teoría de las GANs.
 - Construcción de GANs.
 - Autoencoders y Variational Auto Encoders

- 8 Despliegue en producción:
 - Transfer Learning
 - Estándares de transferencia de modelos ONNX
 - Desplegar en Docker containers.
 - Entrenar usando GPUs y TPUs.
 - Desplegar a la nube (GoogleColab, AzureML y Kaggle)

Prácticas estructuradas (10 horas semanales, que en el régimen hemisemestral se impartirán en un régimen intensivo) de :

- Basicos (se anotan horas en cada uno, total 30 horas) :
 - Herramientas básicas para Deep Life Sciences (2)
 - Introducción a MoleculeNet (2)
 - Molecular Fingerprints (2)
 - Creando modelos con TensorFlow (2)
 - Introduction a las Convoluciones para Grafos (2)
 - Molecular Featurizations (2)
 - Interpretability de modelos (2)
 - Entrenamiento Avanzado de Modelos (2)
 - Unsupervised Embeddings para Moléculas (2)
 - Predicción de Ki de Ligandosa una proteína (2)
 - Modelado de interacciones proteína-ligando, con u sin atomic convolutions.(2)

Fecha	V.01
Página 3 de 6		

Deep Learning aplicado al diseño de compuestos bioactivos
Posgrado

- Entrenando una Generative Adversarial Network (4)
- Chemical Screens en grandes volúmenes de datos (4).
- Reproducir experimentos (cada uno 10 horas total 30 horas) :
 - Ramsundar, Bharath, et al. "Is multitask deep learning practical for pharma?." *Journal of chemical information and modeling* 57.8 (2017): 2068–2076.
 - Delaney, John S. "ESOL: estimating aqueous solubility directly from molecular structure." *Journal of chemical information and computer sciences* 44.3 (2004): 1000–1005.
 - Ma, Junshui, et al. "Deep neural nets as a method for quantitative structure–activity relationships." *Journal of chemical information and modeling* 55.2 (2015): 263–274.
- Trabajo sobre el datasets (cada uno 10 horas total 30 horas):
 - Tox21: Ref: B Ramsundar, S Kearnes, P Riley, D Webster, D Konerding, V Pande arXiv preprint arXiv:1502.02072
 - Clintox: The Clintox dataset is a collection of "clinical toxicity" datasets that compares drugs approved by the FDA and drugs that have failed clinical trials for toxicity reasons. It contains two classification tasks for 1491 compound
 - Harvard Organic photovoltaic: Lopez, Steven A., et al. "The Harvard organic photovoltaic dataset." *Scientific data* 3.1 (2016): 1–7
- **PROYECTO INTEGRADOR: solo exigible en caso de estudiantes de posgrado a quienes se les orientará a lo largo del curso (30 horas) para que desarrollen y presenten al final un proyecto basado en una hipótesis que el propio estudiante propondrá. Para ello contará con un tutor (alguno de los docentes del curso)**
- **GANANCIA: El curso se gana con la presentación de la resolución de un mínimo de 60% de las practicas estructuradas. Estudiantes de grado: podrán exonerar el examen si resuelven un mínimo de 60% de las prácticas estructuradas. Estudiantes de posgrado: además de poder exonerar el examen con un mínimo de las prácticas estructuradas, se les pedirá exposición oral del Proyecto integrador.**

Bibliografía:

Demystifying Deep Convolutional Neural Networks. Adam Harley

<https://www.cs.ryerson.ca/~aharley/neural-networks/>

Ramsundar, B. (2018). *Molecular machine learning with DeepChem* (Doctoral dissertation, Stanford University).

Haghighatlari, M., Vishwakarma, G., Altarawy, D., Subramanian, R., Kota, B. U., Sonpal, A., ... & Hachmann, J. (2019). Chemml: A machine learning and informatics program package for the analysis, mining, and

Fecha	V.01
	Página 4 de 6	



Deep Learning aplicado al diseño de compuestos bioactivos Posgrado

modeling of chemical and materials data. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, e1458.

Wen, Ming, et al. "Deep-learning-based drug–target interaction prediction." *Journal of proteome research* 16.4 (2017): 1401-1409.

Baskin, I. I. (2020). The power of deep learning to ligand-based novel drug discovery. *Expert Opinion on Drug Discovery*, 1-10.

Lin, E., Lin, C. H., & Lane, H. Y. (2020). Relevant Applications of Generative Adversarial Networks in Drug Design and Discovery: Molecular De Novo Design, Dimensionality Reduction, and De Novo Peptide and Protein Design. *Molecules*, 25(14), 3250.

Colby, S. M., Nuñez, J. R., Hodas, N. O., Corley, C. D., & Renslow, R. R. (2019). Deep learning to generate in silico chemical property libraries and candidate molecules for small molecule identification in complex samples. *Analytical chemistry*, 92(2), 1720-1729.

Senior, Andrew W., et al. "Improved protein structure prediction using potentials from deep learning." *Nature* 577.7792 (2020): 706-710.

Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., & Vandergheynst, P. (2017). Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4), 18-42.

ODENA, Augustus. Semi-supervised learning with generative adversarial networks. *arXiv preprint arXiv:1606.01583*, 2016.

GUI, Jie, et al. A review on generative adversarial networks: Algorithms, theory, and applications. *arXiv preprint arXiv:2001.06937*, 2020.

Frameworks:

Keras: https://keras.io/getting_started/

Tensorflow: <https://www.tensorflow.org/learn>

Scikit-learn: https://scikit-learn.org/stable/user_guide.html

Modalidad del Curso:

	Teórico	Practico	Laboratorio	Otros (*)
--	---------	----------	-------------	-----------

Fecha	V.01
	Página 5 de 6	

Deep Learning aplicado al diseño de compuestos bioactivos
Posgrado

Asistencia Obligatoria	4	2		
Modalidad Flexible	100%	100%		

Régimen de ganancia:

La nota del curso se compone de dos partes para estudiantes de grado: i) prácticos semanales ii) entregable final. Los porcentajes de cada componente para la nota final es el siguiente: 60% prácticos semanales, y 40% entregable final.

La nota del curso se compone de tres partes para estudiantes de posgrado: i) prácticos semanales ii) entregable final iii) exposición oral del Proyecto Integrador. Los porcentajes de cada componente para la nota final es el siguiente: 60% prácticos semanales, y 20% entregable final. 20% Proyecto Integrador

La nota de la asignatura (nota final) corresponde 100% de la nota del curso.

Requisitos para ganar el curso: el estudiante debe alcanzar un 66% en su calificación.

Fecha	V.01
		Página 6 de 6