UDELAR FQ FACULTAD DE QUIMICA	**Deep Learning Aplicado al Diseño de Compuestos Bioactivos 2024** **POSGRADO**

Carácter del curso	**POSGRADO**
Semestre en que se dicta	Segundo Hemisemestre Semestre PAR
Número de créditos	13
Carga horaria semanal (hs)	6 (teóricos) y 10 (practicas estructuradas). Proyecto Integrador: 30 horas
Previaturas	Programación Python I(Inicial) y/o Programación Python II (Intermedio)
Cupo	30

Estructura Responsable:
Área Bioinformática, DETEMA, Facultad de Química, UdelaR.

Docente Responsable:
Margot Paulino
Pablo García

Docentes colaboradores
Jorge Cantero
Andrés Camilo Ballesteros

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.

UDELAR FQ FACULTAD DE QUIMICA	**Deep Learning Aplicado al Diseño de Compuestos Bioactivos 2024** **POSGRADO**

Contenido:

Clases teóricas de 2 horas cada una

1 Introducción al Diseño de Compuestos Bioactivos.

2 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.

3 Regularización y optimización (hyperparameters tuning).

Métodos de evaluación (Mean Squared Error, Binary Cross Entropy, Categorical Cross Entropy)
Hyperparámeters tuning.

Inicialización
Regularización Dropout.
Normalización batch.

4 Métodos de optimización:

Gradient Descent
Stochastic Gradient Descent
Momentum
Variable and Adaptive Learning Rates
Adam

5 Convolutional Neural Networks :

Concepto de convolución y kernels comúnmente utilizados.
Max Polling
Aplicación de las redes convolucionales a estructuras 3D e imágenes.

6 Redes neuronales deep, otras arquitecturas de uso específico:

RNN Recurrent Neural Networks
LSTM Long Short Term Memory

7 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

8 GAN Generative adversarial networks

Modelos generativos.
Teoría de las GANs. Construcción de GANs.
Autoencoders y Variational Auto Encoders

9 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)

TOTAL TEÓRICOS: 16 horas. (4 horas por semana)

TOTAL Prácticas estructuradas (Total 30 horas para grado)

Desarrollo de las Prácticas Estructuradas Básicas (se anotan horas en cada una, 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)

Entrenando una Generative Adversarial Network (4) Chemical
Screens en grandes volúmenes de datos (4).

PROYECTO INTEGRADOR: 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)

Desarrollo de las 30 horas de actividades de discusión (entregables)

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 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 (*)
Asistencia Obligatoria	4	**2**		
Modalidad Flexible	**100%**	**100%**		

Régimen de ganancia:

El estudiante debe asistir a por lo menos el 80% de la clases del curso, tanto teóricos, teórico-prácticos como las horas destinadas al proyecto integrador.

Las evaluaciones consistirán en un parcial escrito sobre los contenidos teóricos del curso que aportará el 40% de la ganancia y una presentación escrita y defensa sobre el proyecto integrador realizado, que aportará el restante 60%. El estudiante que obtenga, entre ambas evaluaciones, al menos 60% del puntaje, exonerará el curso, de lo contrario, deberá dar examen.

Por mayor información visitar la página del curso o consultar directamente con el docente responsable margot@fq.edu.uy