

EL7024 Teoría de Información: Fundamentos y Aplicaciones**Profesor:** Jorge F. Silva**Auxiliar:** Mauricio González**Ayudantes:** Felipe Córdova, Sebastián Espinosa, Mario Vicuña y Miguel Videla

Proyecto de Curso

1 Marco General

El objetivo del proyecto es entender y familiarizarse algunos conceptos aprendidos en el curso EL7024 mediante la investigación de un problema que utilice herramientas de teoría de la información. El proyecto debe considerar alguno de los siguientes aspectos:

- Uso o estudio de herramientas de teoría de información en un contexto aplicado.
- Investigar un tema específico de teoría de información o aprendizaje de máquinas que involucre una componente de validación experimental.

El proyecto se espera que sea de 2 personas y consta de las siguientes etapas:

- Primera entrega: 2 de abril; Definición del tema a investigar presentando 3 proyectos ordenados por prioridad.
- Segunda entrega: 5 de Junio; Entrega de progresos consistente en un Informe de avance. El informe debe tener al menos las secciones de abstract, introducción, desarrollo y resultados preliminares.
- **Examen (fecha por definir):** Entrega Final: Entrega de Informe final y presentación oral (15 minutos). El informe debe tener al menos las secciones de abstract, introducción, desarrollo, resultados y conclusiones.

El formato del informe debe ser tipo paper (puede ser en Word o Latex), se subirá a U-Cursos una plantilla de informe.

2 Propuestas de Proyecto

Propuestos por el cuerpo docente.

2.1 Análisis teórico de redes neuronales y deep learning

Tutor: Mario Vicuña

Si bien las técnicas de aprendizaje profundo han significado un éxito casi rotundo en una gran cantidad de tareas de inferencia, de la mano de un avance constante en aspectos técnicos (nuevas arquitecturas, mayor poder de cómputo, etc) una gran incógnita que queda por resolver son los fundamentos teóricos detrás de este tipo de algoritmos de aprendizaje. En esta línea el método del Information Bottleneck ha surgido como una forma de análisis del problema [1], [2], ofreciendo relaciones con el principio de máxima verosimilitud, pero también como una metodología para justificar y entender el uso de ciertos regularizadores explícitos e implícitos. En la literatura hay una gama de trabajos que buscan dar una justificación a este principio, y también visiones críticas al respecto [3], [4]. Por esto, se invita a los estudiantes a revisar y entender las diferentes propuestas, compararlas teórica y/o experimentalmente y, a partir de esto, generar su propio punto de vista crítico, ya sea avalando alguna perspectiva en particular o proponiendo análisis propios y novedosos.

Referencias recomendadas:

- 1 Tishby, N. and Zaslavsky, N., *Deep Learning And The Information Bottleneck Principle*. IEEE Information Theory Workshop. 2015
- 2 Schwartz-Ziv, R. and Tishby, N., *Opening The Black Box Of Deep Neural Networks Via Information*. arXiv preprint arXiv:1703.00810. 2017
- 3 Saxe, A., Bansal, Y., Dapello, J., Advani, M., Kolchinsky, A., Tracey, B. and Cox, D., *On the Information Bottleneck Theory of Deep Learning*. Journal of Statistical Mechanics: Theory and Experiment. 2018

- 4 Amjad, R. and Geiger, B., *Learning Representations For Neural Network-Based Classification Using The Information Bottleneck Principle*. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2019
- 5 Hafez-Kolahi, H. and Kasaei, S., *Information Bottleneck And Its Applications In Deep Learning*. arXiv preprint arXiv:1904.03743. 2019

2.2 Incorporación de ruido y estocasticidad para mejorar el aprendizaje

Tutor: Mario Vicuña

Desde hace bastante tiempo se asume que añadir ruido a arquitecturas profundas permite lograr una mejor generalización a partir de datos supervisados usados en entrenamiento. Esto en principio puede verse como poco intuitivo pero existen resultados que permiten justificar este principio. El caso más clásico en redes profundas es el dropout que incluso a día de hoy es una de las técnicas de regularización más usadas y aceptadas en el área. Más avanzadas técnicas se han explorado, por ejemplo el dropout variacional [1] que aprovecha localidades en los datos para añadir ruido en forma adecuada. En este contexto, el principio del Information Bottleneck ha sido explorado como una alternativa para estas técnicas, siendo un caso ilustre el trabajo que propone el Information dropout [2]. Por lo tanto, se invita a los estudiantes a explorar estas técnicas desde una perspectiva del Information Bottleneck, a niveles tanto teóricos como prácticos. Se invita a realizar comparaciones o desarrollar alguna variante original sustentada en este principio, como es el caso de [3] en Reinforcement Learning.

Referencias recomendadas:

- 1 Kingma, D., Salimans, T. and Welling, M., *Variational Dropout And The Local Reparameterization Trick*. Advances in neural information processing systems. 2015
- 2 Achille, A. and Soatto, S., *Information Dropout: Learning Optimal Representations Through Noisy Computation*. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2018
- 3 Igl, M., Ciosek, K., Li, Y., Tschitschek, S., Zhang, C., Devlin, S. and Hofmann, K., *Generalization in Reinforcement Learning with Selective Noise Injection and Information Bottleneck*. Conference on Neural Information Processing Systems. 2019

2.3 Codificación universal

Tutor: Sebastián Espinosa

El objetivo de este proyecto es realizar un análisis bibliográfico de compresión cuando no se conocen las estadísticas, en particular, se pueden explorar los diferentes algoritmos y sus garantías de optimalidad (LZ77, LZ78). Se debe explorar los conceptos de redundancia y regret y comparar los desempeños de los algoritmos con la mejor tasa alcanzable. En esta línea se sugiere como aplicación la clasificación de textos mediante el uso de herramientas de codificación universal. Para esto se debe utilizar conceptos aprendizaje y compresión cuando no se conocen las estadísticas.

Referencias recomendadas:

- 1 L. Davisson, *Universal Noiseless Coding*, IEEE Transactions on Information Theory. 1973
- 2 Ryabko, *Information Theoretic Method for Classification of Texts*, Problems of Information Transmission. 2017
- 3 Cover & Thomas, *Elements of Information Theory*, John Wiley & Son. 2012

2.4 Aplicaciones del *Information Bottleneck*

Tutor: Mario Vicuña

Basados en el principio del *Information Bottleneck*, existen arquitecturas de aprendizaje que se han modificado para incorporar estos principios de Teoría de la Información, como es el caso de InfoGAN o su variante IB-GAN [1]. Otras ideas corresponden a [2] y [3] en las cuales se busca restringir el flujo de información en manera "inteligente", en estos casos aplicado en GANs y problemas de aprendizaje reforzado. Desde el ámbito del Machine Learning se invita a los estudiantes a seguir esta línea de llevar los conceptos detrás del Information Bottleneck a los algoritmos de aprendizaje y estudiar cómo estos afectan en su performance, comparando y analizando distintas arquitecturas (pudiendo incluir arquitecturas propias basadas en este principio). No obstante, cabe destacar que este proyecto no se limita solo al aprendizaje de máquinas, dados los orígenes del Information Bottleneck en el ámbito de la codificación y comunicaciones, este posee

naturalmente una gama de aplicaciones en estas áreas [4], las cuales los estudiantes están invitados a abordar, tanto a nivel teórico como práctico, incluyendo un análisis crítico de el/los problemas estudiados bajo este principio.

Referencias recomendadas:

- 1 Jeon, I., Lee, W. and Kim, G., 2019. *IB-GAN: Disentangled Representation Learning With Information Bottleneck GAN*.
- 2 Peng, X., Kanazawa, A., Toyer, S., Abbeel, P. and Levine, S., 2019. *Variational Discriminator Bottleneck: Improving Imitation Learning, Inverse RL, and GANs by Constraining Information Flow*. International Conference on Learning Representations.
- 3 Goyal, A., Islam, R., Strouse, D., Ahmed, Z., Botvinick, M., Larochelle, H., Bengio, Y. and Levine, S., 2019. *InfoBot: Transfer and Exploration via the Information Bottleneck*. International Conference on Learning Representations.
- 4 Zaidi, A., Aguerri, I. and Shamaï, S., 2020. *On The Information Bottleneck Problems: Models, Connections, Applications And Information Theoretic Views*. Entropy

2.5 Estimación de la información mutua con particiones adaptativas y no adaptativas del espacio de observación

Tutor: Mauricio González

En diversas aplicaciones se requiere estimar la información mutua entre dos distribuciones desconocidas a partir de muestras de las mismas (realizaciones independientes e idénticamente distribuidas). Para ello una alternativa natural es utilizar estimadores empíricos, cuya clase más directa son los estimadores *plug-in*. Estos estimadores primero estiman las distribuciones a partir de los datos (es decir, se obtienen distribuciones empíricas) y luego se utilizan éstas para calcular la información mutua, considerando que la información mutua es un funcional del espacio de distribuciones.

El objetivo de este proyecto es evaluar el desempeño de estimadores *plug in* estimando las distribuciones al usar particiones no adaptativas del espacio (tipo producto, independiente de los datos) y adaptativas (donde las celdas se concentran donde hay más datos).

Además, se espera que puedan profundizar en aspectos formales del problema como es el estudio de consistencia y también comparar las velocidades de convergencia mediante un análisis longitudinal en un escenario experimental adecuado.

Referencias recomendadas:

- 1 G. Darbellay and I. Vajda, *Estimation of the Information by an Adaptive Partitioning of the Observation Space*, IEEE Transactions on Information Theory, vol. 45, no. 4, May 1999
- 2 J. Silva and S. Narayanan, *Non-Product Data-Dependent Partitions for Mutual Information Estimation: Strong Consistency and Applications*, IEEE Transactions on Signal Processing, 2010.

2.6 Problemas de estimadores *plug in* para símbolos infrecuentes

Tutor: Mauricio González

La entropía es una medida de información de una variable aleatoria. Considerando el caso discreto, sea $X \in B$ una variable aleatoria con distribución de probabilidad de masa P_X . Su entropía está dada por:

$$H(X) = H(P_X) = \sum_{x \in B} -P_X(x) \log(P_X(x))$$

Al usar el estimador *plug in*, se tiene que:

$$\widehat{H(X)} = H(\widehat{P_X}) = \sum_{x \in B} -\widehat{P_X}(x) \log(\widehat{P_X}(x)) = \sum_{x \in B} f(\widehat{P_X}(x)),$$

donde $f(t) = t \log_2(1/t)$. El problema surge cuando tenemos símbolos infrecuentes y $\widehat{P_X}(x) \approx 0$, pues en ese caso $f'(t)$ es grande (tendiendo a infinito cuando t tiende a 0 por la derecha). Luego, pequeños errores en la estimación de la probabilidad empírica llevan a grandes errores en el funcional de entropía $\widehat{H(X)}$.

Así, el objetivo de este proyecto es investigar estrategias para abordar este problema. Un enfoque recomendado es separar $f(t)$ reemplazando su porción inicial con un polinomio, de manera de evitar la porción con derivada elevada. Implementar el nuevo enfoque y compararlo con el estimador *plug in* clásico mediante un diseño experimental adecuado.

Referencias recomendadas:

- 1 J. Jiao, K. Venkat, Y. Han and T. Weissman, *Minimax Estimation of Functionals of Discrete Distributions*, IEEE Transactions on Information Theory, 2015.
- 2 Y. Wu and P. Yang, *Minimax rates of entropy estimation on large alphabets via best polynomial approximation*, IEEE Transactions on Information Theory, February 2016.

2.7 Test de independencia utilizando medidas de información

Tutor: Mauricio González

La información mutua es una medida de información que cuantifica la dependencia estadística entre dos variables aleatorias, de tal manera que dadas las v.a. X, Y , se tiene que $I(X; Y) = 0$ si y solo si X e Y son independientes. El proyecto considera el escenario en que las distribuciones de ambas variables aleatorias son desconocidas, y que se dispone de muestras de ambas. El objetivo de este proyecto es proponer un detector de independencia (test de hipótesis binario) por medio del uso de estimadores empíricos de la información mutua. Se espera que se trabaje en la formalización del problema y en ver la forma de adoptar estimadores. Analizar la consistencia del test y validarlo experimentalmente sería dimensiones deseables.

Referencias recomendadas:

- 1 A. Gretton and L. Györfi, *Consistent Nonparametric Tests of Independence*, Journal of Machine Learning Research, vol. 11, April 2010.
- 2 T. Barret and R. Samworth, *Nonparametric Independence testing via mutual information*, Statistical Laborator, November 2017.

2.8 Redes GAN en test de independencia condicional

Tutor: Felipe Córdova

El estudio de dependencia entre dos variables aleatorias siempre ha sido de interés para la comunidad científica. Un foco interesante de uso es verificar que las representaciones y características obtenidas por redes neuronales sean independientes entre si. Naturalmente el test de independencia no es un problema trivial pues involucra caracterizar atributos estructurales de distribuciones de probabilidad a partir de las observaciones que se disponen. En este contexto las populares GAN permiten aprender distribuciones de probabilidad a partir de datos con muy buena precisión (expresividad de los modelos) y por lo tanto es posible utilizar este hecho para abordar el problema de test de independencia.

El trabajo que motiva este proyecto [1] utiliza redes GAN condicionales para realizar test de independencia condicional para muestras (X, Y, Z) donde se busca verificar si $P(X, Y|Z) = P(X|Z)P(Y|Z)$, es decir, si la variable X es independiente a la variable Y cuando se condicionan con respecto a Z . El trabajo se basa en que la red GAN condicional aprenda la distribución $p_{X|Z}$ de los datos para luego generar datos sintéticos \tilde{X} y con estos simples generados y los datos disponibles crear un test de hipótesis que bajo un estadístico permite determinar si las variables son independientes o no con un cierto grado de confiabilidad asociado al error de tipo I en el setting de test de hipótesis.

El proyecto se basa en estudiar los conceptos teóricos que se presentan en [1] y replicar los test de independencia que allí se presentan. Como se utiliza el entrenamiento de una red neuronal condicional, se propone aplicar a la red propuesta en [1] la regla de regularización presentada en InfoGan [2] que busca mejorar la preservación de información de la variable condicional. Un aspecto interesante a estudiar es que el enfoque presentado en [2], para la regularización, depende de estimadores de información mutua, por tanto, se propone estudiar métodos tradicionales de información mutua en conjunto con el estimador MINE [3], que se cubrirá en unas de las cápsulas del curso.

Referencias recomendadas:

- 1 A. Bellot and M. van der Schaar, *Conditional independence testing using generative adversarial networks*. Advances in Neural Information Processing Systems, 2019.

- 2 X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Ab-beel, *Infogan: Interpretable representation learning by information maximizing generative adversarial nets*. Advances in neural information processing systems, 2016.
- 3 I. Belghazi, S. Rajeswar, A. Baratin, R. D. Hjelm, and A. C. Courville, *MINE: mutual information neural estimation*. arXiv preprint arXiv:1801.04062. 2018.

2.9 Corrección de Loss para el entrenamiento de la red generadora en redes GAN

Tutor: Felipe Córdova

La función objetivo clásica de las redes GAN se puede escribir como:

$$\min_G \max_D \mathbb{E}[\log D(x)] + \mathbb{E}[\log(1 - D(G(z)))]$$

Este problema puede ser interpretado como el clásico adversarial game en donde la red Discriminadora se encarga de diferenciar si los datos que recibe provienen de la distribución original o son generados virtualmente (por un agente adversario), mientras que la red Generadora trata de engañar a la red Discriminadora generando datos sintéticos. Otro enfoque es visualizar el problema como la minimización de la divergencia como se plantea de forma mas general en el trabajo sobre f -GAN [2], en donde para un caso particular de la función f definida se puede recuperar la función objetivo clásica.

Como es de esperar, la optimización de estas 2 redes se realiza por pasos en donde para la función discriminadora los 2 términos que aparecen en el criterio de optimización se utilizan para propagar el gradiente a los pesos de la red, mientras que la red generadora solo considera el segundo termino para estos propósitos. El objetivo de este proyecto se basa en el trabajo presentado en [1] en donde se explora el hecho que la optimización que realiza la red Generadora no corresponde a una divergencia y por tanto se pierde el foco de optimalidad buscado, en donde se trata imitar la distribución original de los datos.

En este proyecto, se espera el estudio de conceptos teóricos que cuestionen el criterio de optimización de las redes GAN y propongan formas de re-pensarlo para alcanzar mayor optimalidad. Además se pide estudiar las funciones objetivo propuesta, donde sí se considera una optimización de una divergencia (sobre la red generadora) la función d debe cumplir que $d(x, y) \geq 0$ y $d(x, y) = 0$ ssi $x = y$. Las comparaciones de los modelos propuestos en [1] pueden basarse en las distintas formas de plantear el problema en [2].

Referencias recomendadas:

- 1 A. Jolicoeur-Martineau, *Gans beyond divergence minimization*. arXiv preprint arXiv:1809.02145, 2018.
- 2 S. Nowozin, B. Cseke, and R. Tomioka, *f-gan: Training generative neural samplers using variational divergence minimization*. Advances in neural information processing systems 2016.

2.10 Aprendizaje de representaciones profundas mediante maximización de información mutua

Tutor: Felipe Córdova

Uno de los objetivos principales del aprendizaje profundo consiste en obtener representaciones de utilidad de las entradas de la red para la realización de diversas tareas (representation learning). En este ámbito, las medidas de información resultan adecuadas para obtener representaciones relevantes. Uno de los más recientes modelos diseñados para esta tarea es el denominado *Deep InfoMax* [1]. Este método consiste en un modelo de aprendizaje no supervisado que maximiza la información mutua entre las entradas y la salida de una red neuronal profunda (codificadora). Para este fin se usan estimadores de información mutua (parametrizados por redes neuronales), pudiendo adicionalmente controlar las características de la representación obtenida ajustando una distribución a priori de manera adversaria.

El proyecto consiste en el estudio, implementación y experimentación con el modelo *Deep InfoMax*, incorporando distintos estimadores actuales de información mutua.

Referencias recomendadas:

- 1 Hjelm, R. D., Fedorov, A., Lavoie-Marchildon, S., Grewal, K., Bachman, P., Trischler, A., Bengio, Y. *Learning deep representations by mutual information estimation and maximization*. arXiv preprint arXiv:1808.06670, 2018.

- 2 Belghazi, M. I., Baratin, A., Rajeswar, S., Ozair, S., Bengio, Y., Courville, A., Hjelm, R. D. *Mine: mutual information neural estimation*. arXiv preprint arXiv:1801.04062. 2018
- 3 Nowozin, S., Cseke, B., Tomioka, R. *f-gan: Training generative neural samplers using variational divergence minimization*. In Advances in neural information processing systems (pp. 271-279). 2016
- 4 Oord, A. V. D., Li, Y., Vinyals, O. *Representation learning with contrastive predictive coding*. arXiv preprint arXiv:1807.03748. 2018
- 5 Liangjian Wen, Yiji Zhou, Lirong He, Mingyuan Zhou, Zenglin Xu. *Mutual Information Gradient Estimation for Representation Learning*, In International Conference on Learning Representations. 2020

2.11 Estimadores de información mutua basados en cotas variacionales

Tutor: Miguel Videla

La estimación de la información mutua se ha tornado esencial en una gran cantidad de problemas en aprendizaje de máquinas, sin embargo, muchos de estos estimadores presentan serios problemas frente al escalamiento en la cantidad y la dimensionalidad de los datos. Para solventar los problemas de escalabilidad y tratabilidad, se ha estudiado una diversidad de estimadores de información mutua basados en cotas variacionales parametrizados por redes neuronales.

El proyecto consiste en el estudio, implementación y experimentación con diversos estimadores de información mutua basados en cotas variacionales, buscando encontrar similitudes teóricas, limitaciones, además de ventajas y desventajas frente a distintos escenarios experimentales.

Referencias recomendadas:

- 1 Poole, B., Ozair, S., Oord, A. V. D., Alemi, A. A., Tucker, G. *On variational bounds of mutual information*. arXiv preprint arXiv:1905.06922. 2019
- 2 Belghazi, M. I., Baratin, A., Rajeswar, S., Ozair, S., Bengio, Y., Courville, A., Hjelm, R. D. *Mine: mutual information neural estimation*. arXiv preprint arXiv:1801.04062. 2018
- 3 Nowozin, S., Cseke, B., Tomioka, R. *f-gan: Training generative neural samplers using variational divergence minimization*. In Advances in neural information processing systems (pp. 271-279). 2016
- 4 Oord, A. V. D., Li, Y., Vinyals, O. *Representation learning with contrastive predictive coding*. arXiv preprint arXiv:1807.03748.
- 5 Nguyen, X., Wainwright, M. J., Jordan, M. I. *Estimating divergence functionals and the likelihood ratio by convex risk minimization*. IEEE Transactions on Information Theory, 56(11), 5847-5861. 2010

2.12 Estimación de gradiente de información mutua para aprendizaje de representaciones

Tutor: Miguel Videla

La estimación de información mutua resulta de alta relevancia en el aprendizaje de representaciones, sin embargo, esta resulta usualmente intratable en escenarios continuos o de alta dimensionalidad. En la actualidad se han propuesto una diversidad de estimadores basados en cotas variacionales y parametrizaciones mediante redes neuronales capaces de solventar dicho problema, no obstante, la mayoría de estos estimadores no son capaces de proveer una estimación precisa y de baja varianza de la información mutua cuando esta alcanza grandes valores. Recientemente se ha propuesto un método de estimación directa del gradiente de la información mutua [1] basada en la estimación de una función de puntaje de distribuciones implícitas, evitando estimar directamente la información mutua como tal, mostrando una ostensible mejora de desempeño y precisión en diversas áreas de aplicación.

El proyecto consiste en el estudio teórico e implementación del estimador de gradiente de información mutua, explorando distintos ámbitos de aplicación y contrastando su desempeño en las distintas tareas frente a respectivos métodos del estado del arte.

Referencias recomendadas:

- 1 Liangjian Wen, Yiji Zhou, Lirong He, Mingyuan Zhou, Zenglin Xu, *Mutual Information Gradient Estimation for Representation Learning*, In International Conference on Learning Representations. 2020
- 2 Li, Y., Turner, R. E. *Gradient estimators for implicit models*. arXiv preprint arXiv:1705.07107. 2017

- 3 Belghazi, M. I., Baratin, A., Rajeswar, S., Ozair, S., Bengio, Y., Courville, A., Hjelm, R. D. *Mine: mutual information neural estimation*. arXiv preprint arXiv:1801.04062. 2018
- 4 Hjelm, R. D., Fedorov, A., Lavoie-Marchildon, S., Grewal, K., Bachman, P., Trischler, A., Bengio, Y. *Learning deep representations by mutual information estimation and maximization*. arXiv preprint arXiv:1808.06670. 2018
- 5 Hlynsson, H. D., Wiskott, L. *Learning gradient-based ICA by neurally estimating mutual information*. In Joint German/Austrian Conference on Artificial Intelligence (Künstliche Intelligenz) (pp. 182-187). Springer, Cham. 2019

2.13 Función de costo robusta a ruido de etiquetas basada en medidas de información para entrenamiento de redes neuronales

Tutor: Miguel Videla

El etiquetado de grandes bases de datos para entrenar algoritmos de aprendizaje supervisado resulta altamente costoso por lo que se prefiere adquirir etiquetado de baja calidad, lo que induce a etiquetados erróneos que impactan negativamente el desempeño de los modelos entrenados. En este contexto, se han desarrollado una gran variedad de métodos para el aprendizaje sobre etiquetas ruidosas los cuales, usualmente, carecen de justificación teórica o requieren de información adicional sobre los datos. Recientemente se propuso una función de costo basada en métricas de información robusto al ruido de etiquetas [1], la cual puede ser aplicada a cualquier red neuronal de clasificación de manera directa sin requerir de ninguna información adicional, cuya efectividad se encuentra teóricamente justificada.

El proyecto consiste en el estudio, implementación, aplicación de funciones de costo robustos a ruido de etiqueta, con énfasis en la función de costo basada en métricas de información anteriormente mencionada, realizando un análisis comparativo de las limitaciones, fortalezas y debilidades de los métodos en distintos escenarios experimentales.

Referencias recomendadas:

- 1 Xu, Y., Cao, P., Kong, Y., Wang, Y. . LDML: A Novel *Information-theoretic Loss Function for Training Deep Nets Robust to Label Noise*. In Advances in Neural Information Processing Systems (pp. 6222-6233). 2019
- 2 Ghosh, A., Kumar, H., Sastry, P. S. *Robust loss functions under label noise for deep neural networks*. In Thirty-First AAAI Conference on Artificial Intelligence. 2017
- 3 Kong, Y. *Dominantly Truthful Multi-task Peer Prediction with a Constant Number of Tasks*. In Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (pp. 2398-2411). Society for Industrial and Applied Mathematics. 2020

2.14 Fotometría Bayesiana

Tutor: Sebastián Espinosa

El objetivo es caracterizar y formalizar el problema de estimación Bayesiana para el flujo de un objeto puntual luminoso (estrella). Se debe comparar el desempeño de estimadores usados en este contexto (por ejemplo la esperanza condicional o la regla maxima a posteriori) con la cota de Crámer-Rao Bayesiana (o cota de Van-Trees). Para esto último se pueden simular los datos a partir del modelo estadístico asociado a éste problema y luego comparar el desempeño empirico con la cota teórica.

Referencias recomendadas:

- 1 Echeverria, A., Silva, J. F., Mendez, R. A., and Orchard, M *Analysis of the Bayesian Cramér-Rao lower bound in astrometry Studying the impact of prior information in the location of an object*, Astronomy & Astrophysics. 2016

2.15 Invarianzas en aprendizaje

Tutor: Sebastián Espinosa

El objetivo de este proyecto consiste en describir y caracterizar el error de generalización en clasificadores invariante. En particular, se considera el escenario donde la clasificación es invariante frente a ciertas transformaciones de la entrada (por ejemplo, rotaciones o traslaciones en una imagen). Luego el clasificador aprende a ser invariante a dichas transformaciones. Se espera que el estudiante formalice la noción de una transformación invariante, cómo repercute en el error

de generalización (ganancia) y simular algún escenario práctico y simple.

Referencias recomendadas:

- 1 J. Sokolíc, R. Giryes, G. Sapiro, M. Rodrigues *Generalization Error of Invariant Classifiers*, arXiv preprint arXiv:1610.04574. 2016
- 2 J. Sokolíc, R. Giryes, G. Sapiro, M. Rodrigues *Robust Large Margin Deep Neural Networks*, IEEE Transactions on Signal Processing. 2017
- 3 H. Xu, S. Mannor *Robustness and generalization*, Machine Learning. 2012

2.16 Otros ejemplos de temas

Propuesto por los estudiantes (previa autorización del cuerpo docente).

Sitios de interés:

- [Entropy Open Access Journal](#).
- [IEEE Xplore](#).
- [Conference on Neural Information Processing Systems \(NeurIPS\)](#).
- [International Conference on Learning Representations \(ICLR\)](#).