

# *Abstract*

## **Advanced Kernel Methods for Multi-Task Learning**

by Carlos Ruiz Pastor

directed by José R. Dorronsoro Ibero and Carlos M. Alaíz Gudín

Machine Learning (ML), whose goal is to automatize the process of learning, has a great influence in our current society. The ML algorithms try to infer general pattern from data, which can then be applicable to new unseen data. These algorithms, such as the Support Vector Machine (SVM) or the Neural Network (NN), are present in many daily situations and have strongly impacted multiple areas such as engineering or advertising, among many others. Multi-Task Learning (MTL) is a field of ML that considers learning different tasks jointly to improve the learning process. This is the natural way of learning for humans: we do not learn each task in an isolated manner, but there exist related tasks that are better learned together. The goal in MTL is to develop strategies to mimic this behaviour, where learning various tasks jointly offers an advantage.

This thesis begins with a presentation of some basic concepts and definitions that we will use in the rest of this work. After that, a theoretical motivation for MTL is given, and some of the most relevant works in this area are reviewed. A taxonomy for these methods is proposed, where three categories are considered: feature-based, parameter-based and combination-based strategies. Different ML algorithms, depending on their characteristics, are more suitable for one of these strategies than for others. The feature-based strategies are more natural for NNs, while the kernel methods, such as the SVM, present a more rigid framework, and the combination-based strategies are better suited for them.

Considering the combination-based strategies, a new convex formulation is proposed: in each task we consider a convex combination of a common and task-specific part as the model. This formulation offers some nice properties, such as better interpretability of the models, or the possibility to dismiss the common or task-specific parts with a particular choice of the hyperparameters. This approach is applied to kernel methods, in particular, the L1, L2 and LS-SVM with this convex MTL formulation are proposed, and the solutions for the corresponding training problems can be obtained with standard SVM solvers. One natural alternative, which considers the direct convex combination of pre-trained common and task-specific models, is also described. In multiple experiments, it is observed that the kernel methods with a convex MTL formulation obtain better results than those considering just a common model, task-specific ones or the convex combination of pre-trained models. As a real world application, the prediction of solar and wind energy is also presented using these models, where our proposal outperforms or ties with the described competition.

Applying this formulation, an MTL proposal for NNs is also made, where a convex combination of a common and a task-specific networks is used. These models can also be trained with standard optimization techniques for NNs. In experiments with four image datasets, it is shown that the results of this proposal are better than using standard approaches, such as sharing the weights in the hidden layers and task-specific output neurons.

Another approach is proposed with a Graph Laplacian (GL) regularization, where the tasks are interpreted as nodes in a graph, and the pairwise distances between the task models are penalized. In this approach, the adjacency matrix of the graph defines the weights for the distances. A new formulation, based on the tensor product of Reproducing Kernel Hilbert Spaces, to apply this regularization in kernel spaces is developed, and the GL regularization is applied to the L1, L2 and LS-SVM. It is exemplified with multiple experiments that this approach can obtain competitive results. Moreover, an algorithm to automatically learn the graph adjacency matrix from the data is proposed and examples of the advantages of using this algorithm are given using experiments with synthetic and real data.

The thesis ends with some general conclusions and presents lines of research for future work.

# *Resumen*

El aprendizaje automático (AA), cuyo objetivo es automatizar el proceso de aprendizaje, tiene una gran influencia en la sociedad actual. Los algoritmos de AA tratan de inferir patrones generales a partir de datos, los cuales pueden después ser aplicados a nuevos datos. Estos algoritmos, como las Máquinas de Vectores Soporte (MVS) o las Redes Neuronales (RN), están presentes en muchas situaciones cotidianas y han causado un fuerte impacto en múltiples áreas como la ingeniería o la publicidad, entre otras muchas. El Aprendizaje Multitarea (AMT) es un campo del AA que considera el aprendizaje de diferentes tareas de forma conjunta para mejorar el proceso de aprendizaje. Esta es la manera natural de aprender para los humanos: no aprendemos las tareas de manera aislada, sino que existen tareas que están relacionadas y se aprenden mejor juntas. El objetivo del AMT es desarrollar estrategias que imiten este comportamiento, donde aprender diversas tareas conjuntamente ofrece una ventaja.

Esta tesis comienza presentando algunos conceptos básicos y definiciones que usaremos en el resto de este trabajo. Después de esto, se ofrece una motivación teórica para el AMT, y se revisan algunos de los trabajos más relevantes en este área. Se propone una taxonomía para estos métodos, donde se consideran tres categorías: estrategias basadas en características, basadas en parámetros y basadas en combinación. Distintos algoritmos de AA, dependiendo de sus características, son más aptos para una u otra de estas estrategias. Las estrategias basadas en características son naturales para las RN, mientras que los métodos de kernel, como las MVS, presentan un esquema más rígido, y son más adaptables a las basadas en combinaciones.

Dentro de las estrategias basadas en combinaciones, se presenta una nueva formulación convexa: se considera la combinación convexa de una parte común y otra específica como el modelo de cada tarea. Esta formulación ofrece algunas propiedades buenas, como una mejor interpretabilidad o la posibilidad de, con una selección particular de hiperparámetros, eliminar la parte común o la específica del modelo. Este enfoque se aplica a métodos de kernel, en particular se proponen la L1, L2 y LS-MVS convexas para AMT, y las soluciones correspondientes a los problemas de entrenamiento se pueden obtener con técnicas estándares de MVS. También se describe una alternativa natural, que considera la combinación convexa directa de un modelo común y otros específicos que ya han sido preentrenados. Se observa en múltiples experimentos que los métodos de kernel con esta formulación convexa obtienen mejores resultados que considerando únicamente un modelo común, uno específico, o la combinación de estos modelos preentrenados. También se presenta la predicción de energía eólica y solar como una aplicación real de estos modelos, donde nuestra propuesta iguala o supera a otros enfoques, como el de un modelo común, modelos específicos por tarea o la combinación convexa directa.

Aplicando esta formulación se hace una propuesta de AMT para RN, donde se considera una combinación convexa de redes comunes y específicas. Estos modelos se puede entrenar también con técnicas estándares para RN. Se muestra con cuatro conjuntos de imágenes que los resultados de nuestra propuesta son mejores que aquellos obtenidos con enfoques más tradicionales, como compartir las capas ocultas y definir neuronas de salida específicas para cada tarea.

Se propone además otro enfoque con una regularización basada en el laplaciano de un grafo, en el que las tareas se interpretan como nodos en un grafo, y las distancias entre los modelos de cada par de tareas son penalizadas. En este enfoque, la matriz de adyacencia del grafo define el peso de cada distancia. Se propone una nueva formulación, basada en el producto tensorial de espacios de Hilbert con kernel reproductor, para usar este enfoque, y la regularización laplaciana se aplica a la L1, L2 y LS-MVS. Se ejemplifica con múltiples experimentos que este enfoque puede obtener resultados competitivos. Además se propone un algoritmo para aprender la matriz de adyacencia de forma automática y se proporcionan ejemplos de las ventajas de este algoritmo en varios problemas sintéticos y reales.

Esta tesis termina con algunas conclusiones generales y presenta líneas de investigación para un trabajo futuro.