

Resumen

Esta tesis se enmarca en la intersección entre las técnicas modernas de Machine Learning, como las Redes Neuronales Profundas, y el modelado probabilístico confiable. En muchas aplicaciones, no solo nos importa la predicción hecha por un modelo (por ejemplo *esta imagen de pulmón presenta cáncer*) sino también la confianza que tiene el modelo para hacer esta predicción (por ejemplo *esta imagen de pulmón presenta cáncer con 67% probabilidad*). En tales aplicaciones, el modelo ayuda al tomador de decisiones (en este caso un médico) a tomar la decisión final. Como consecuencia, es necesario que las probabilidades proporcionadas por un modelo reflejen las proporciones reales presentes en el conjunto al que se ha asignado dichas probabilidades; de lo contrario, el modelo es inútil en la práctica. Cuando esto sucede, decimos que un modelo está perfectamente calibrado.

La regla de decisión de Bayes proporciona un marco fundamentado para la toma de decisiones en condiciones de incertidumbre y garantiza un rendimiento óptimo (es decir, probabilidades mínimas de error). Para que funcione la regla de decisión de Bayes, es necesario utilizar un modelo calibrado, ya que eso implica que el modelo ha recuperado (aproximado mejor) la distribución de generación de datos. La calibración no es lo único que importa, sino también el refinamiento que es la capacidad del clasificador para recuperar cómo se separan los datos de las diferentes clases.

Sin embargo, las técnicas modernas de aprendizaje automático, como las redes neuronales profundas, no están calibradas, lo que compromete su implementación en aplicaciones de alto riesgo. Numerosos trabajos han intentado

solucionar el error de calibración de las redes neuronales profundas modernas, y este es uno de los principales objetivos de esta tesis.

Esta tesis comienza revisando los elementos involucrados en la regla de decisión de Bayes, a través de las denominadas *Proper Scoring Rules*. Esto nos permite introducir uno de los conceptos clave, al menos en mi opinión personal, que se debe tener en cuenta para entrenar un modelo de aprendizaje automático. Esta es la incertidumbre de los datos en la distribución objetivo, es decir, cómo se superponen las diferentes muestras de las diferentes clases. Esta observación se utiliza en la primera contribución de esta tesis para justificar por qué cualquier técnica de aumento de datos no garantizará distribuciones calibradas, a pesar de que la evidencia empírica se ha proporcionado en la dirección opuesta. Mostramos cómo una función de pérdida propuesta que tiene en cuenta la incertidumbre de los datos resuelve el error de calibración introducido por el entrenamiento Mixup, una técnica de aumento de datos de última generación.

Sin embargo, dado que las redes neuronales profundas son modelos costosos de entrenar, las técnicas que tienen como objetivo calibrar implícitamente estos modelos son costosas de implementar en la práctica, ya que uno tiene que lidiar con técnicas de selección de modelos. Para ello, la segunda contribución propone recalibrar la salida de una Red Neural Profunda usando una Red Neural Bayesiana. Con esto, mostramos que se pueden usar modelos expresivos siempre que se incorpore la incertidumbre, lo que contrasta con muchas de las contribuciones recientes que plantean la hipótesis de que el espacio de calibración es inherentemente simple porque las técnicas más simples funcionan mejor que sus contrapartes complejas. También mostramos que la principal crítica de las técnicas bayesianas, cuando se aplican a las redes neuronales modernas, se supera combinando las capacidades de las redes neuronales profundas con la red neuronal bayesiana desacoplada propuesta.

Uno de los problemas de las redes neuronales bayesianas es la especificación de un prior significativo en el espacio de parámetros que induce un prior útil en el espacio funcional. Un priori mal especificado sesgará erróneamente la forma en que cuantificamos la incertidumbre con el posterior, lo que conducirá a predicciones bayesianas subóptimas. En la contribución final de esta tesis, presentamos un nuevo previo que se aplica directamente al espacio funcional, denominado *Transformed Gaussian Process*. Este nuevo prior sobre el espacio de funciones se construye deformando muestras de un proceso gaussiano utilizando una transformación invertible. Estas funciones de deformación están parametrizadas por redes neuronales bayesianas, que nos permiten modelar procesos no estacionarios que tienen en cuenta la incertidumbre de los parámetros, mejorando claramente el rendimiento sobre la estimación puntual. Intro-

ducimos un algoritmo de inferencia variacional escasa que nos permite aligerar la carga computacional que heredaríamos de los procesos gaussianos estándar, aproximar al posterior intratable, entrenar el modelo usando inferencia variacional estocástica y usar cualquier modelo de observación; entre otras propiedades.