

Resum

Aquesta tesi s'emmarca en la intersecció entre les tècniques modernes d'aprenentatge automàtic, com les xarxes neuronals profundes, i el modelatge probabilístic fiable. En moltes aplicacions, no només ens preocupa la predicció feta per un model (*aquesta imatge pulmonar presenta càncer*), sinó també la confiança del model per fer aquesta predicció (*aquesta imatge pulmonar presenta càncer amb probabilitat de 67 %*). En aquestes aplicacions, el model ajuda el que pren la decisió (en aquest cas un metge) a prendre la decisió final. Com a conseqüència, cal que les probabilitats que proporciona un model reflecteixin el veritable conjunt subjacent de resultats, en cas contrari, el model no serà inútil a la pràctica. Quan això passa, diem que un model està perfectament calibrat.

La regla de decisió de Bayes proporciona un marc de principis per a la presa de decisions sota incertesa i garanteix un rendiment òptim (és a dir, probabilitats mínimes d'error). Perquè la regla de decisió de Bayes funcioni, cal utilitzar un model calibrat, ja que això implica que el model ha recuperat (millor) la distribució que genera dades. La calibració no és l'únic que importa, sinó també el refinament que és la capacitat del classificador per recuperar la forma en què es separen les dades de les diferents classes.

Tot i això, les tècniques modernes d'aprenentatge automàtic, com les xarxes neuronals profundes, no estan calibrades, cosa que compromet el seu desplegament en aplicacions d'alt risc. Molts treballs han intentat resoldre l'error de calibració de les xarxes neuronals profundes modernes, i aquest és un dels principals objectius d'aquesta tesi.

Aquesta tesi comença revisant els elements implicats en la regla de decisió de Bayes, a través de la lent de les regles de puntuació adequades. Això ens permet introduir un dels conceptes clau, almenys en la meua opinió personal, que s'hauria de tenir en compte per formar un model d'aprenentatge automàtic. Aquesta és la incertesa de les dades en la distribució objectiu, és a dir, com se superposen les diferents mostres de les diferents classes. Aquesta observació s'utilitza en la primera contribució d'aquesta tesi per justificar per què qualsevol tècnica d'augment de dades no garantirà distribucions calibrades, tot i que s'han aportat proves empíriques en la direcció contrària. Mostrem com una proposta de funció de pèrdua que té en compte la incertesa de les dades resol el mal calibrat introduït per l'entrenament Mixup, una tècnica d'augment de dades d'última generació.

Tanmateix, atès que les xarxes neuronals profundes són models cars per entrenar, les tècniques que volen calibrar implícitament aquests models costen de desplegar-se a la pràctica, ja que s'ha de fer front a tècniques de selecció de models. Per a això, la segona contribució proposa recalibrar la sortida d'una xarxa neuronal profunda mitjançant una xarxa neuronal bayesiana. Amb això, demostrem que es poden utilitzar models expressius sempre que s'incorpori la incertesa, cosa que contrasta amb moltes de les recents contribucions que hipoteten que l'espai de calibratge és intrínsecament senzill perquè les tècniques més senzilles funcionen millor que les seves contraparts complexes. També demostrem que les principals crítiques a les tècniques bayesianes, quan s'apliquen a les xarxes neuronals modernes, es superen combinant les capacitats de les xarxes neuronals profundes amb la xarxa neuronal bayesiana desacoblada proposada.

Un dels problemes de les xarxes neuronals bayesianes és l'especificació d'un prior significatiu a l'espai de paràmetres que indueix un prior útil a l'espai de funcions. Un anterior especificat malament esbiaixarà erròniament la manera de quantificar la incertesa amb el posterior, la qual cosa conduirà a prediccions bayesianes subòptimes. En la contribució final d'aquesta tesi, introduïm un nou prior que s'aplica directament a l'espai de funcions, anomenat Procés Gaussià Transformat. Aquest nou prior sobre funcions es construeix mitjançant la deformació de mostres d'un procés gaussià mitjançant una transformació inversible. Aquestes funcions de deformació estan parametritzades per les xarxes neuronals bayesianes, que ens permeten modelar processos no estacionaris que tinguin en compte la incertesa dels paràmetres, millorant clarament el rendiment respecte a la contrapart estimada per punts. Introduïm un algorisme d'inferència variacional escassa que ens permet alleugerir la càrrega computacional que heretaríem dels processos gaussians estàndard, orientar-nos

a la part posterior intractable, entrenar el model mitjançant la inferència variacional estocàstica i utilitzar qualsevol model d'observació; entre altres boniques propietats.