

Optimization methods for risk scores learning

Student: Luca Ciabini, 2021

Blaylock Risk Assessment Screening Score (BRASS) index - (Blaylock 1992)
Età <i>(una sola opzione)</i> 0 = 55 anni o meno 1 = 56 – 64 anni 2 = 65 – 79 anni 3 = 80 anni e più
Condizioni di vita e supporto sociale <i>(una sola opzione)</i> 0 = Vive col coniuge 1 = Vive con la famiglia 2 = Vive solo con il sostegno del familiare 3 = Vive solo con il sostegno di amici/conoscenti 4 = Vive solo senza alcun sostegno 5 = Assistenza domiciliare/residenziale
Stato funzionale <i>(ogni opzione valutata)</i> 0 = Autonomo (indipendente in ADL e IADL) <u>Dipendente in:</u> 1 = Alimentazione/nutrizione 1 = Igiene/abbigliamento 1 = Andare in bagno 1 = Spostamenti/mobilità 1 = Incontinenza intestinale 1 = Incontinenza urinaria 1 = Preparazione del cibo 1 = Responsabilità nell'uso di medicinali 1 = Capacità di gestire il denaro 1 = Fare acquisti 1 = Utilizzo di mezzi di trasporto
Stato cognitivo <i>(una sola opzione)</i> 0 = Orientato 1 = Disorientato in alcune sfere* qualche volta 2 = Disorientato in alcune sfere* sempre 3 = Disorientato in tutte le sfere* qualche volta 4 = Disorientato in tutte le sfere* sempre 5 = Comatoso <i>*sfere: spazio, tempo, luogo e sé</i>

Risk scores are classification models that predict the risk of an event using a value defined by the sum of a few integers. They are particularly interesting as they are easy

to learn, to use, to validate and, besides, are transparent. In this thesis we present some improvements to the LCPA algorithm used to solve the RiskSlim problem and create optimized risk scores.

In particular we study feature selection and continuous features handling techniques, trying to embed a partial one-hot encoding in the optimization problem.

Experimental results show how the combination of feature selection and continuous features optimal split brings improvements to the model in terms of quality without affecting too much the learning time.