

Inleiding

De gegevens

Stel dat we een gegevensset hebben, waarbij elk punt een aantal inputs heeft en daarenboven tot een bepaalde klasse of groep hoort. Deze inputs kunnen bijvoorbeeld verschillende eigenschappen zijn van een tumor. De tumor is ofwel kwaadaardig, ofwel goedaardig. De aard van de tumor zal dan de klasse zijn waartoe een datapunt behoort.

Predictie

We willen nu, gegeven een nieuwe dataset, graag voorspellen of een tumor al dan niet kwaadaardig is. Hiervoor kunnen we een machine learning techniek gebruiken die zo'n voorspelling doet voor de classificatie van een tumor met bepaalde eigenschappen. Hiervoor bestaan verschillende technieken, maar indien de hoeveelheid eigenschappen van de tumoren - wat dus meer algemeen overeenkomt met de inputs van de datapunten, zal blijken dat Support Vector Machines hier een zeer geschikte techniek voor blijken te zijn.

Inhoud van dit eindverslag

We zullen in dit eindverslag de lezer eerst laten ontdekken wat machine learning precies is. Vervolgens bespreken we wat supervised learning inhoudt, in vergelijking met unsupervised learning. We bespreken ook wat classificatie juist is. Hierna zullen we het principe achter Support Vector Machines, de toepassing waarvoor we kozen, uitleggen. Hierbij zullen we de verschillende classificatietechnieken vergelijken en beargumenteren waarom voor de tumor-dataset SVM een geschikte keuze blijkt te zijn.