

Variantie in het kader van machine learning is de maat voor de gevoeligheid van een model. Het geeft inzicht over de flexibiliteit van het model, met name hoe goed het model zich kan aanpassen aan verschillende datasets. De variantie analyseert het verschil tussen de door het model voorspelde waarde en de werkelijke waarde van een variabele. Voor het berekenen van de variantie wordt het verschil tussen de werkelijke waarde en de voorspelde waarde genomen en dit vervolgens gekwadrateerd. Neem een variabele x , dan is $f(x)$ de waarde horende bij de variabele x en $f(\hat{x})$ de voorspelde waarde voor x . Dan kan de variantie als volgt worden gedefinieerd:

$$Var(x) = (E(f(x) - f(\hat{x}))^2$$

Om een goed model te verkrijgen is het wenselijk om de variantie zo laag mogelijk te nemen. Een lage variantie betekent dat het model weinig afhankelijk is van veranderingen in de trainingsset en minder gevoelig is voor uitzonderlijke waarden binnen de set. Hierdoor zullen er accuratere voorspellingen gedaan worden bij andere datasets. Bij een hoge variantie past het model zich nauwkeurig aan de trainingsset aan, maar zal het model de eventuele uitschieters ook als belangrijk beschouwen. Wanneer de trainingsset verandert, zal het model dus ook duidelijke aanpassingen vertonen. Bij het gebruik van een andere dataset zal er dan een duidelijker verschil zijn tussen de werkelijke en de voorspelde waarden. Een toename van de variantie zal leiden tot een verminderde nauwkeurigheid van het model.