



Inschatting biedingsstrategie

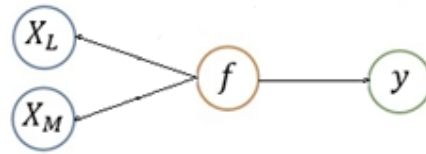
In deze leaflet leggen wij graag het model achter de module ‘inschatting biedingsstrategie’ uit. Het model is een ‘gemixt kansdichtheid netwerk’, ‘Mixture Density Network’ in het Engels. In dit artikel wordt deze term stap voor stap uitgelegd. Elke stap wordt geïllustreerd met een versimpeld voorbeeld. Het doel van dit voorbeeld is uiteindelijk om de kansverdeling van de transactieprijs van een bepaald huis uit te rekenen.

In dit versimpelde voorbeeld heeft het huis maar 2 kenmerken, een locatie-kenmerk en een ‘krapte van de markt’-kenmerk. Het locatie-kenmerk is een continue waarde tussen 0 en 1, waar 0 de minst gewilde locatie is en 1 de meest gewilde locatie voor een huis in Nederland. Het ‘krapte van de markt’-kenmerk is 0 of 1. Een 0 geeft een niet krappe huizenmarkt aan, de meeste huizen worden onder de vraagprijs verkocht. Een 1 geeft een krappe huizenmarkt aan, de meeste huizen worden boven de vraagprijs verkocht. In de rest van dit artikel, de notatie van deze kenmerken is X_L and X_M , voor respectievelijk het locatie-kenmerk en het markt-kenmerk. De transactieprijs wordt beschreven door y .

Een gemixt kansdichtheid netwerk is een regressie-model dat gebruik maakt van een neuraal netwerkarchitectuur. Graag leggen we hieronder eerst kort de termen regressie en neuraal netwerk uit. Ook wordt het concept kansdichtheid plot toegelicht. Regressie, een neuraal netwerk en de kansdichtheid zijn samen de ingrediënten voor het recept van een gemixt kansdichtheid netwerk. Hiermee is het model in staat de kans op meerdere mogelijke uitkomsten te voorspellen. Het model kan hierdoor beter omgaan met complexe voorspellingen waarbij onzekerheid een rol speelt.

Regressie

Regressie is het benaderen van een functie f dat het verband vangt tussen kenmerken X en een uitkomst y , in ons voorbeeld:



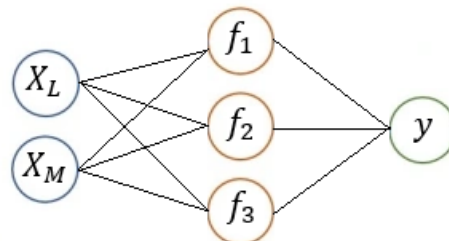
Dit kan een eenvoudige lineaire regressie zijn zoals:

$$y = 300.000 + 400.000 \times X_L + 100.000 \times X_M$$

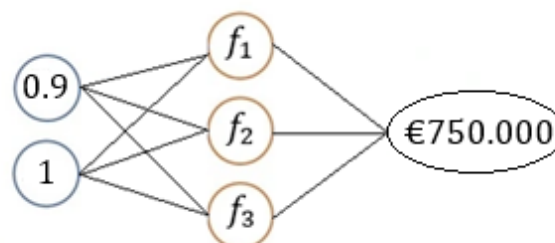
Het bovenstaande regressie-model zou voor een huis op een erg gewilde locatie ($X_L = 0.9$) in een krappe markt ($X_M = 1$) als transactieprijs $300.000 + 400.000 \times 0.9 + 100.000 \times 1 = \text{€}760.000$ voorspellen. Een lineaire regressie is het meest eenvoudige model dat kan worden gebruikt en te simplistisch voor het voorspellen van een bod op een woning. Een beter model is een neurale netwerk, zie de volgende paragraaf.

Neuraal netwerk

In een neurale netwerk wordt het verband tussen de kenmerken X en een uitkomst y gevangen door een complexe non-lineaire functie, details worden hier verder niet gegeven om het niet te ingewikkeld te maken.



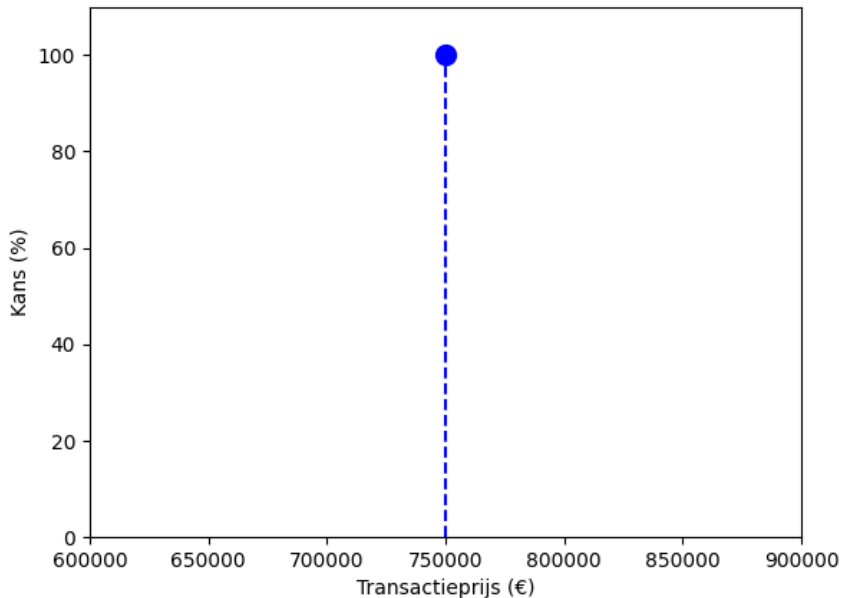
Voor het huis in ons voorbeeld voorspelt het neurale netwerk dat de transactieprijs $\text{€}750.000$ is.



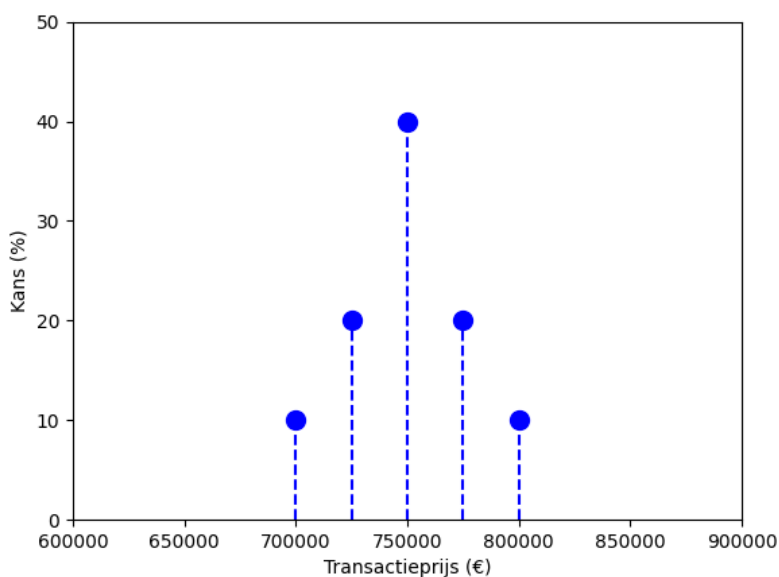
Voor nu is het belangrijk om te begrijpen dat we woning- en locatiekenmerken hebben voor een huis, en dat een complexe functie voor deze kenmerken de transactieprijs van het huis voorspelt.

Kansdichtheid plot

De uitkomst van zowel een lineaire regressie als een neurale netwerk is één waarde, in ons geval één voorspelde transactieprijs. In het voorbeeld van het neurale netwerk voorspelt het model dat dit huis voor €750.000 verkocht zal worden. In andere woorden, de kans is 100% dat dit huis zal worden verkocht voor €750.000, zie onderstaand kansdichtheid plot.

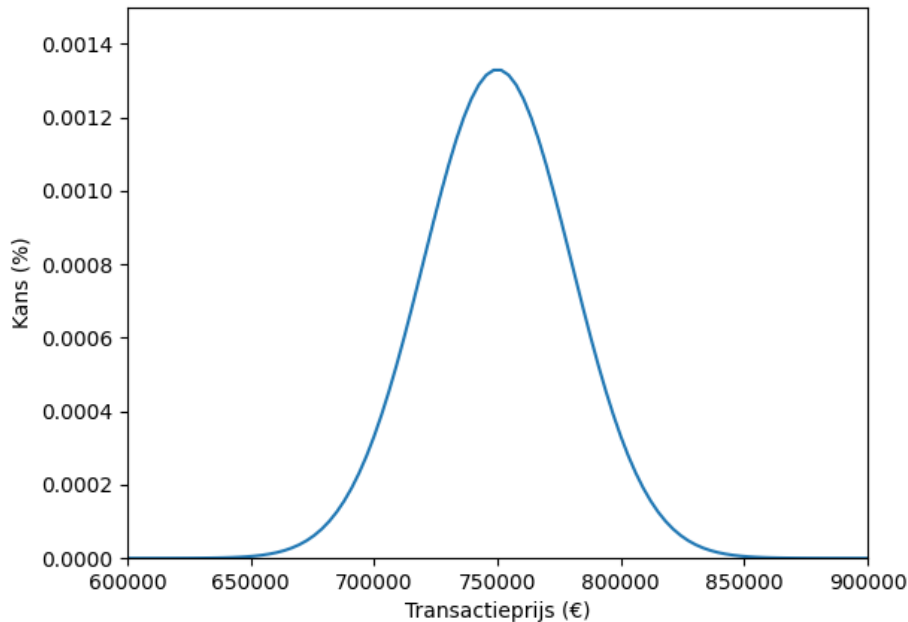


Zo simpel is de werkelijkheid natuurlijk niet. De waarde van 750.000 kan nog zo'n nauwkeurige benadering van de woningwaarde zijn, het sluit niet uit dat koper en verkoper een hoger of lager bedrag overeenkomen. Er zou iemand kunnen zijn die het huis heel graag wilt hebben, en daarvoor €775.000 biedt. Het zou ook zo kunnen zijn dat, voor een bepaalde reden, er juist heel weinig vraag is voor het huis en dat het hoogste bod maar €725.000 is. De kansdichtheid plot ziet er waarschijnlijk meer uit zoals:

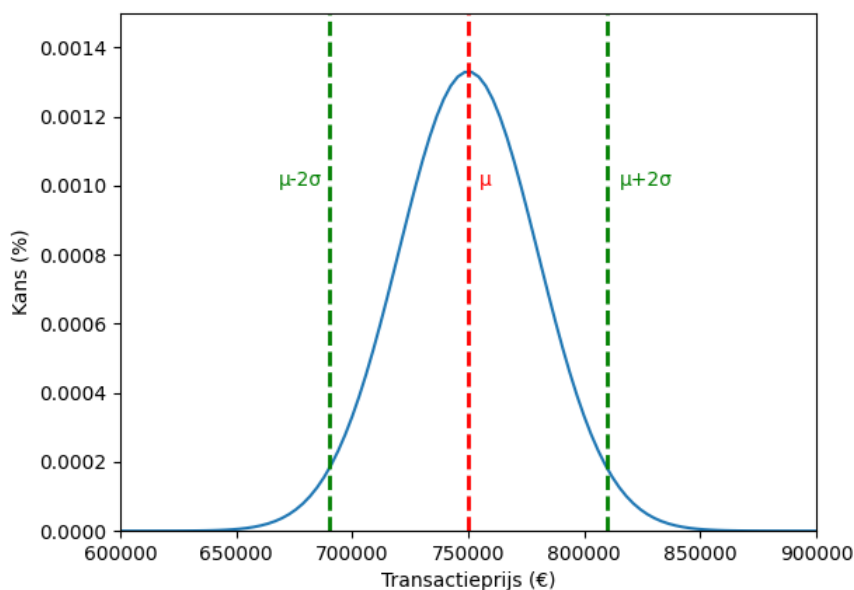


Deze plot vertelt je: met 40% kans zal het huis voor €750.000 verkocht worden, beide met 20% kans zal het huis voor €725.000 of €775.000 worden verkocht en beide met 10% kans zal het huis voor €700.000 of €800.000 worden verkocht.

Belangrijk, merk op dat in beide plots de optelsom van de kansen samen 100% is. Wanneer we de kans uitrekenen voor alle prijzen tussen €600.000 en €900.000, krijgen we een kansdichtheid plot zoals:



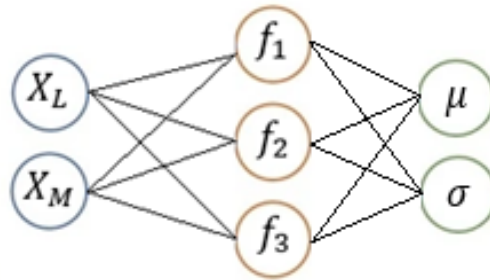
Deze plot ziet er in eerste instantie misschien niet intuïtief uit, de kansen zijn klein maar ook hier tellen ze samen op tot 100%. Bovenstaande plot is gemaakt door een simpele normaalverdeling te plotten. Dit is een kansverdeling die kan worden beschreven met maar 2 parameters, het gemiddelde μ en de standaarddeviatie σ .



Het gemiddelde μ is de waarde die hoort bij de hoogste kans op de curve. De standaarddeviatie σ geeft de breedte weer van de curve.

Kansdichtheid netwerk

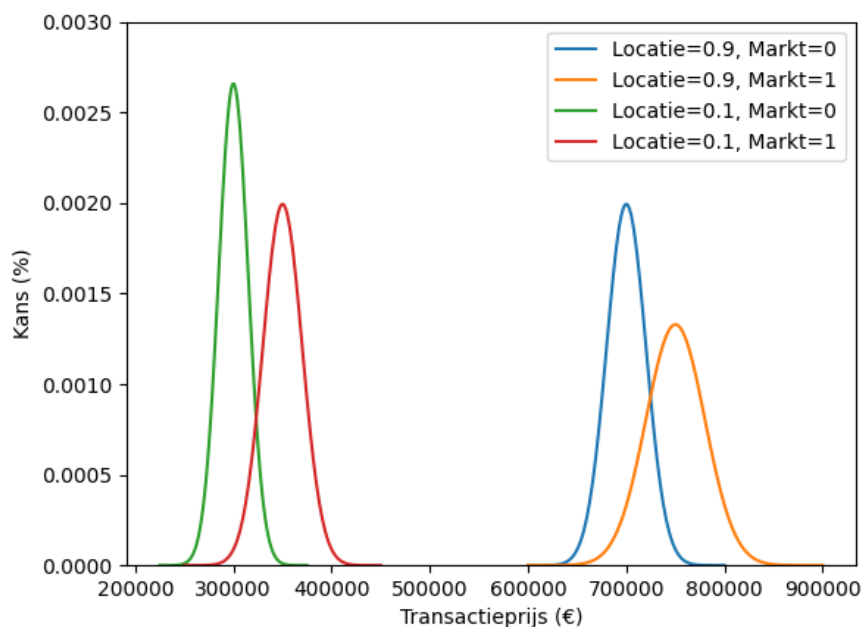
Bij een kansdichtheid netwerk gebruiken we een neurale netwerk om het verband te leren tussen de kenmerken en de parameters van de kansverdeling van de uitkomst, in plaats van de uitkomst zelf. Dit kan worden geïllustreerd door:



In ons voorbeeld zou deze functie de volgende parameters voorspellen voor de verschillende huiskenmerken:

Locatie	Markt	μ	σ
0.9	0	700.000	20.000
0.9	1	750.000	30.000
0.1	0	300.000	15.000
0.1	1	350.000	20.000

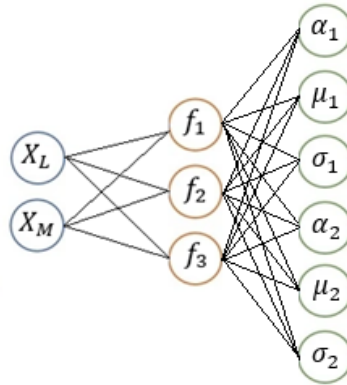
Deze parameters kunnen we gebruiken om de kansdichtheid te plotten voor 4 normaalverdelingen, door de parameters in te vullen in de kansverdeling-formule van een normaalverdeling.



Je ziet duidelijk dat een betere locatie en een krappere markt ertoe leidt dat de verdeling naar rechts opschuift. Je ziet ook dat voor beide huizen (op de mindere en de betere locatie) een krappere markt leidt tot meer variabiliteit op de markt. Dit uit zich in een bredere kansverdeling. Dit zou het resultaat kunnen zijn van door het overbiedproces in een krappere markt.

Gemixt kansdichtheid netwerk

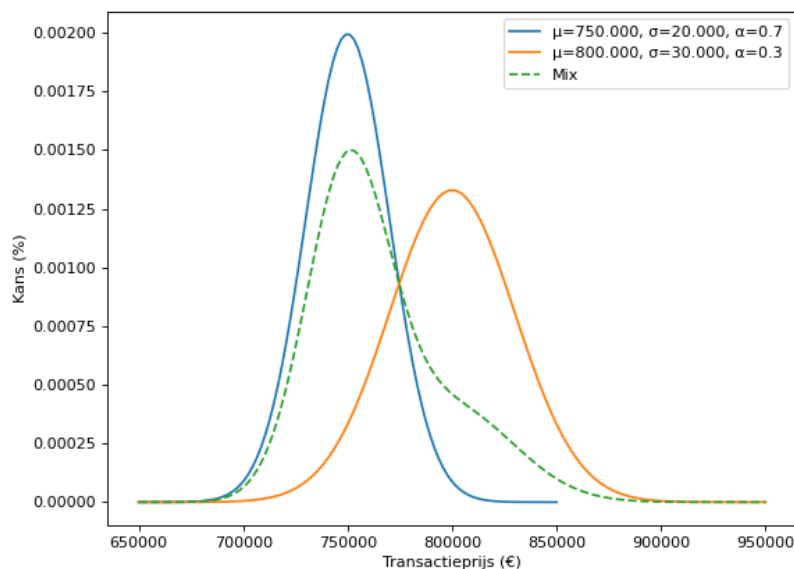
In de praktijk zal de kansdichtheid plot voor elk huis complexer zijn en niet perfect kunnen worden beschreven met een simpele normaalverdeling. Echter, door meerdere normaalverdelingen bij elkaar op te tellen kunnen er al heel wat complexere kansdichtheden worden beschreven. Hieronder geven we een gemixt kansdichtheid netwerk weer, dat het verband uitrekenet tussen de kenmerken en de gewogen som van twee normaalverdelingen:



Als we voor de huiskenmerken $X_L = 0.9$ en $X_M = 1$ invullen, krijgen we de parameters:

Locatie	Markt	α_1	μ_1	σ_1	α_2	μ_2	σ_2
0.9	1	0.7	750.000	20.000	0.3	800.000	30.000

De kansdichtheid-functie die we hieruit krijgen is een mix tussen de eerste normaalverdeling met parameters μ_1 en σ_1 , en de tweede normaalverdeling met parameters μ_2 en σ_2 . De eerste normaalverdeling is meer zichtbaar in de mix, omdat het gewicht α_1 groter is dan het gewicht α_2 . The kansdichtheid plot voor de mix in ons voorbeeld:

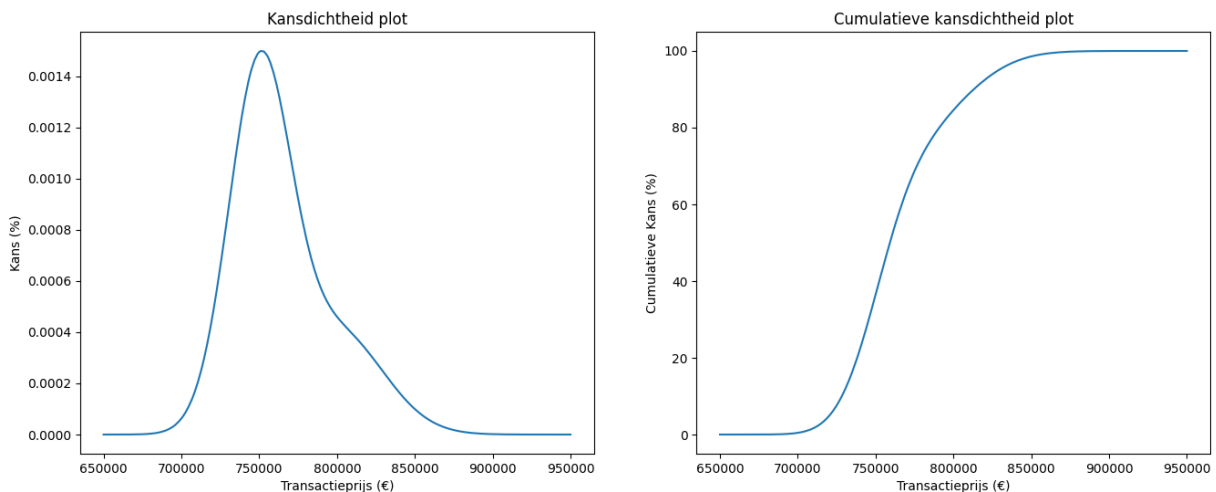


Het is duidelijk dat we curves kunnen definiëren met een meer complexere vorm, bij een mix van meerdere normaalverdelingen. Bij een mix van een groter aantal normaalverdelingen, kunnen we zeer complexe kansdichtheid plots definiëren.

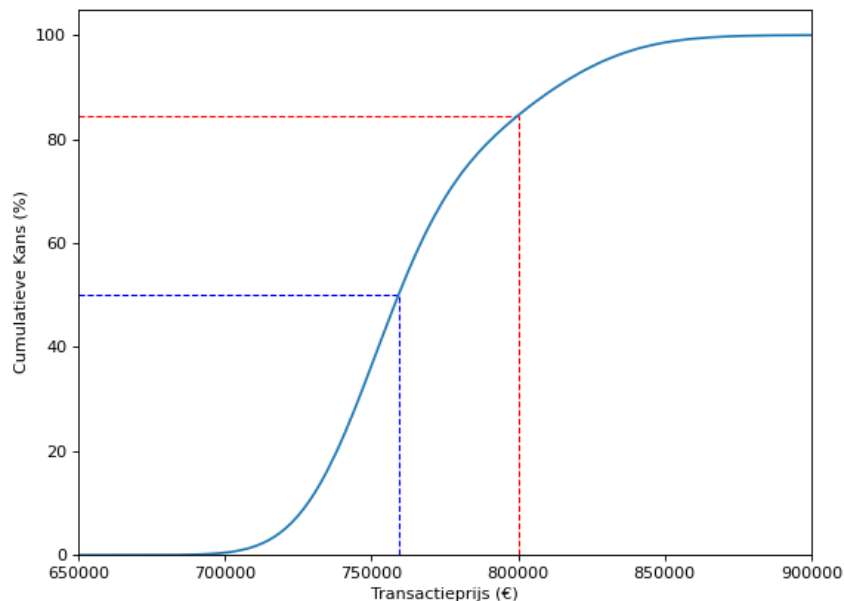
Kortom, een gemixt kansdichtheid netwerk is uitermate geschikt voor het vinden van het verband tussen een grote set kenmerken en de parameters van de kansverdeling van de uitkomst.

Cumulatieve kansdichtheid plot

Er rest ons nog één stap. Het is zeer onwaarschijnlijk dat een potentiële huizenkoper een bod doet dat veel hoger is dan de waarde van de woning. Echter, wanneer dit toch gebeurt, dan is de kans heel hoog dat de verkoper van het huis akkoord gaat met dit bod als uiteindelijke verkoopprijs van de woning. Om dit in kaart te brengen gebruiken wij een *cumulatieve* variant van het kansdichtheid-plot. Voor de inschatting van de biedingsstrategie zijn we geïnteresseerd in de kans dat een huis maximaal voor een bepaalde prijs wordt verkocht. In andere woorden, dat als je een bepaald bod doet, wat is dan de kans dat dit bod het winnende bod is? De kansdichtheid functie kan worden omgevormd naar de cumulatieve kansdichtheid plot door de afzonderlijke kansen bij elkaar op te tellen, daarom de naam cumulatieve kansdichtheid plot.



Met zo'n cumulatieve kansdichtheid plot kan je voor dit voorbeeld zeggen dat als je €800.000 biedt, dat de kans dat dit het winnende bod is ongeveer 84% is. En ook de andere kant op, als je 50% kans wil om het winnende bod te hebben, moet je ongeveer €760.000 bieden.



Voor meer informatie kunt u contact opnemen met support@brainbay.nl of op 030 85004500.