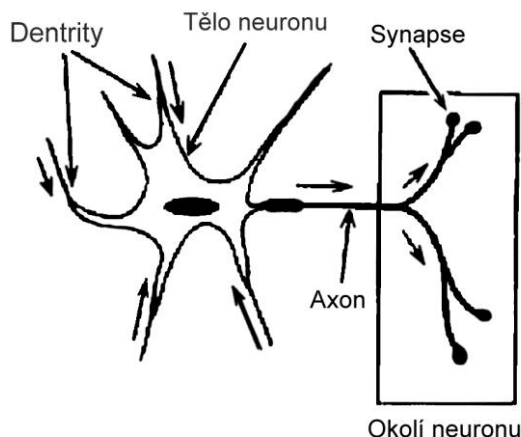


Neuronová síť

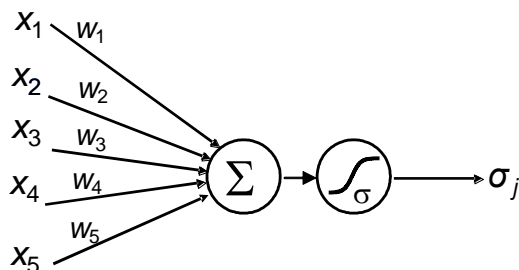
Menu: QCExpert | Prediktivní metody | Neuronová síť

Neuronová síť (Artificial Neural Network, ANN, resp. NN) je velmi populární a výkonná metoda, která se používá k modelování vztahu mezi vícerozměrnou vstupní proměnnou \mathbf{x} a vícerozměrnou výstupní proměnnou \mathbf{y} . Historie neuronových sítí sahá do čtyřicátých let 20. století, avšak výpočetní náročnost příslušných algoritmů umožnila jejich reálné využití až s vývojem výkonných počítačů. NN lze obecně považovat za nelineární regresní model, který lze vyjádřit síťovou strukturou. Inspirací pro neuronové sítě byla struktura mozkové tkáně vyšších živočichů, kde je neuron propojen tzv. synapsí s několika jinými neurony. Synapsí protéká proud (resp. informace, vzruch), který je každým neuronem zpracován a předán dalšími synapsí do dalších neuronů.

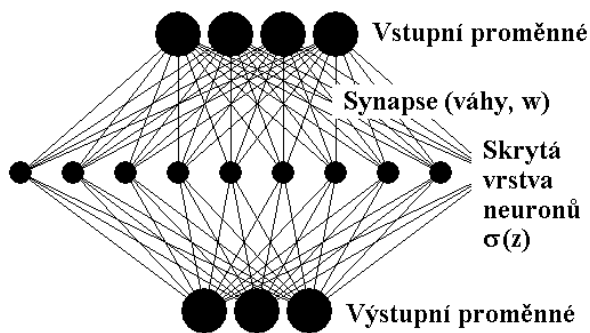


Obrázek 1 Biologický neuron: tělo, vstupní dendrity, výstupní axon, synapse na další neurony

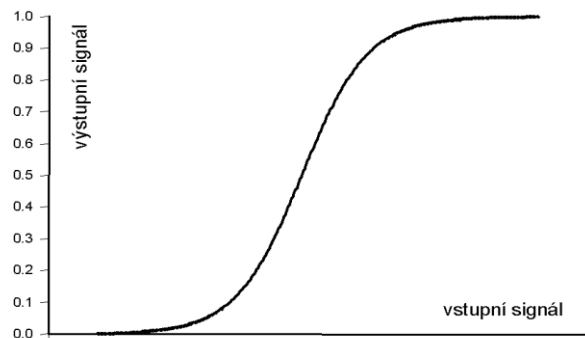
V uzlech neuronové sítě, analogicky nazývaných neurony, je každá vstupní proměnná x_i na vstupu do j -tého neuronu násobena váhovým koeficientem w_{ji} . Součet $z_j = w_{0j} + \sum w_{ji}x_i$ je v neuronu transformován aktivační funkcí. Aktivační funkce vyjadřuje intenzitu odezvy neuronu na daný vstup. Mezi nejběžněji používané aktivační funkce patří logistická funkce, $\sigma_j(z) = 1/(1 + e^{-z})$, která se podobá biologické senzické odezvové funkci. Váhy w_{ji} reprezentují intenzitu vazby mezi proměnnou a neuronem, nebo, v případě vícevrstevných sítí, mezi neurony ve vrstvách, tyto vazby se někdy analogicky nazývají synapse.



Obrázek 2 Schéma umělého neuronu



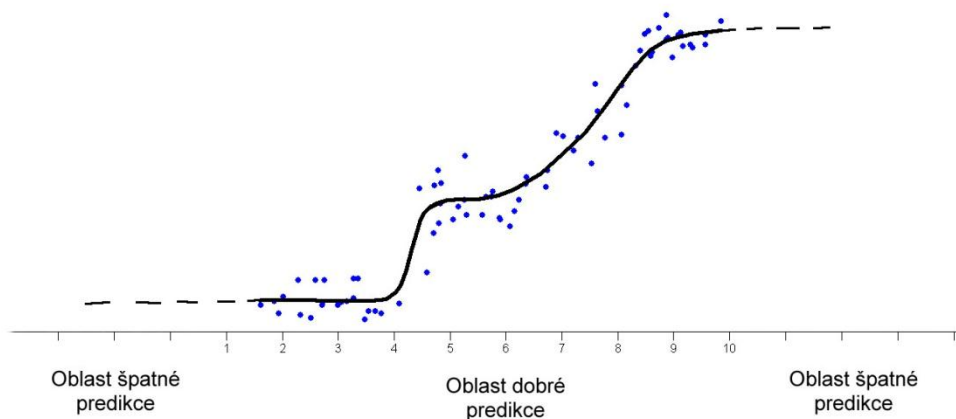
Obrázek 3 Schéma neuronové sítě



Obrázek 4 Aktivační funkce neuronu $\sigma(z)$

Výstupní proměnné se predikují jako vážené lineární kombinace výstupů z poslední skryté vrstvy neuronů, $\hat{y}_i = \sum_k w_{ik} \sigma_{Lk}$. Je tedy neuronová síť formálně zvláštním případem

vícenásobné nelineární regrese, prakticky lze neuronovou síť považovat za neparаметrickou regresi. Kdyby neuronová síť neobsahovala žádnou skrytou vrstvu neuronů – pouze vstupní a výstupní proměnné, jednalo by se o model lineární regrese. Neuronová síť se optimalizuje za kritéria nejmenších čtverců. To znamená, že se síť nastavuje tak, aby součet čtverců rozdílů predikované a naměřené hodnoty výstupní proměnné byl minimální. Toto nastavení je cílem iterativní optimalizační procedury, která se nazývá učení, nebo trénování neuronové sítě. Optimalizační procedura **QCExpertu** využívá adaptivní derivační Gauss-Newtonovské algoritmy. Natrénovaná síť se pak dá využít pro predikci výstupní proměnné, nebo proměnných při zadané kombinaci vstupních proměnných. Model neuronové sítě je lokální, to znamená, že jeho predikční schopnost prudce klesá mimo rozsah zadaných hodnot nezávisle proměnné, viz Obrázek 5.



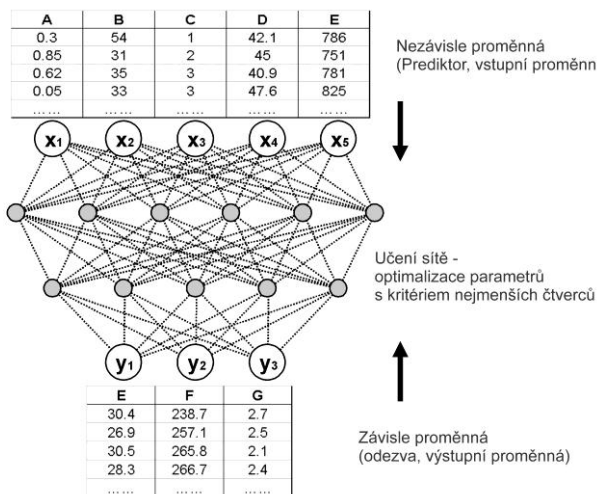
Obrázek 5 Predikční oblasti NN – příklad pro 1-rozměrný prediktor x a 1-rozměrný výstup y

Typický postup použití neuronové sítě je tedy následující.

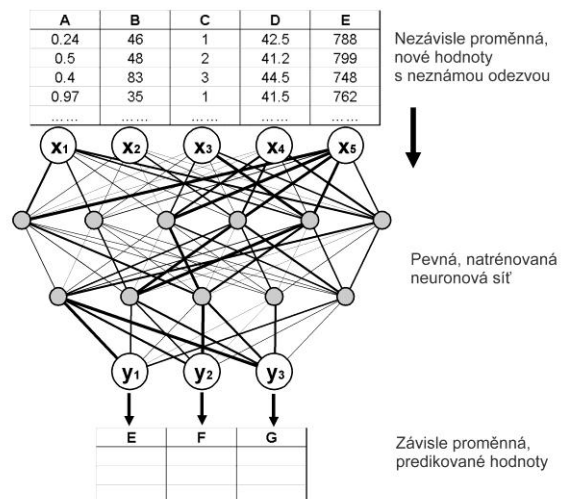
1. Zvolíme skupinu prediktorů – nezávisle proměnných, o kterých se domníváme, že mohou mít vliv na závisle proměnné. Zvolíme skupinu závisle proměnných, na něž mají mít vliv prediktory. V každém řádku musí být vždy hodnoty závisle i nezávisle proměnné, které si odpovídají. Počet řádků označme N .
2. Zvolíme architekturu neuronové sítě – počet vrstev a počty neuronů v jednotlivých vrstvách. Neexistuje jednoznačné pravidlo pro nejlepší architekturu sítě, obvykle je vhodné používat počet neuronů zhruba odpovídající počtu proměnných. Síť s jednou skrytou vrstvou neuronů použijeme tam, kde předpokládáme lineární, nebo slabě nelineární vztahy. Dvouvrstvé sítě modelují silně nelineární vztahy. Používání

vícevrstvých sítí obvykle nebývá efektivnější. Navíc je třeba počítat s tím, že velmi složitá síť mohou být přeuročené, nejednoznačné a nestabilní a lze je jen těžko optimalizovat. Příklady vhodných architektur uvádí Obrázek 18 na str. 7. Počet dat (řádků) by měl být nejméně desetkrát větší, než počet neuronů v síti, jinak roste riziko přeuročení modelu a snížení predikční schopnosti. Obvykle vystačíme i pro rozsáhlé problémy se sítěmi se 2 - 20 neurony a 1 - 3 vrstvami.

3. Optimalizaci parametrů sítě, neboli tzv. "učení" neuronové sítě. Během tohoto procesu se optimalizační algoritmus snaží najít takové nastavení vah, aby síť předpovídala hodnoty v co nejlepším souladu se zadanými závisle proměnnými. Tento soulad se měří součtem čtverců, jedná se tedy o metodu nejmenších čtverců, podobně jako v lineární nebo nelineární regresi. Obecně nemůže být zaručeno, že nalezené řešení je nejlepší možné. Proto se někdy může vyplatit spustit optimalizaci vícekrát a použít model s lepším reziduálním součtem čtverců (RSC). Optimalizace začíná s náhodnými hodnotami vah, je proto přirozené, že každé jednotlivé řešení nalezené optimalizací je zcela odlišné od ostatních. I zcela rozdílné nastavení vah v síti může poskytnout prakticky identickou predikci modelu se shodnými RSC pro zadaná data. Rozdíl se pak ovšem může někdy projevit v kvalitě predikce pro neznámá data.
4. Chceme-li informaci o spolehlivosti predikce, můžeme zvolit pouze tzv. trénovací, nebo učící podmnožinu dat o PN řádcích ($0 < P < 1$), které se použijí pro nalezení modelu a zbytek dat, tedy zbývajících $(1 - P)N$ řádků tzv. testovacích, nebo validačních dat pak použijeme na validaci sítě.
5. Úspěšnost neuronové sítě můžeme posoudit podle průběhu poklesu kritéria součtu čtverců, podle těsnosti proložení v grafech predikce závisle proměnných a podle tloušťky úseček spojujících neurony (tloušťka je úměrná absolutní velikosti příslušné váhy, tedy intenzity toku informace směrem dolů).
6. Predikce: Pokud chceme natrénovanou síť využít ještě pro predikování zatím neznámých hodnot závisle proměnné, zvolíme nezávisle proměnné stejné jako v bodu 1. Pak použijeme natrénovanou síť k predikci neznámých hodnot.



Obrázek 6 Proces učení (trénování) neuronové sítě

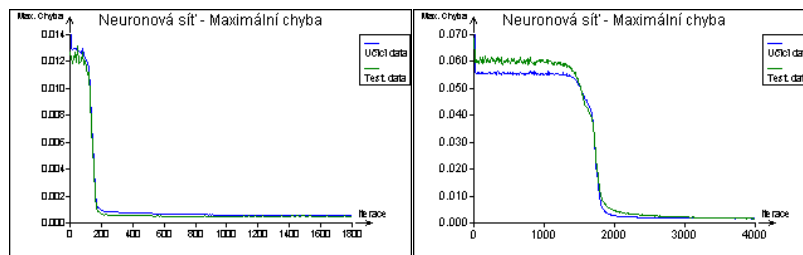


Obrázek 7 Predikce pomocí natrénované neuronové sítě

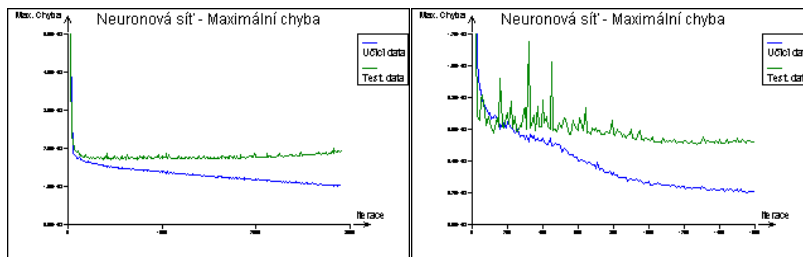
Validace modelu

Modely NN obvykle neumožňují výpočet statistických parametrů a diagnostik pro podrobné posouzení kvality modelu (rozptyly regresních koeficientů, F - a t -statistiky pro

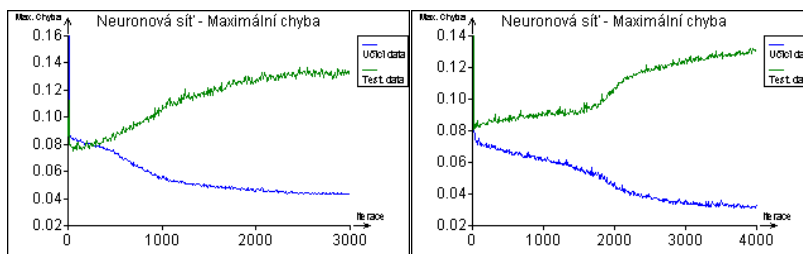
testování významnosti modelu, diagnostické grafy, apod.), jak je tomu například u lineární regrese. Proto je třeba používat jiné metody, abychom se přesvědčili, zda je sestavený model vhodným popisem sledovaného fenoménu. Neuronové sítě jsou velmi flexibilním nástrojem a snadno může dojít k situaci, kdy model bude velmi (až podezřele) dobře popisovat (fitovat) zadaná data, avšak nikoliv fenomén (závislost) jako celek. To se pak projeví velmi špatnou predikcí hodnoty závisle proměnných z nově zadaných hodnot nezávisle proměnné, které se přesně v datech dosud nevyskytly, ačkoliv se nacházejí uvnitř intervalu trénovacích dat.



Obrázek 8 Velmi dobrá predikční schopnost

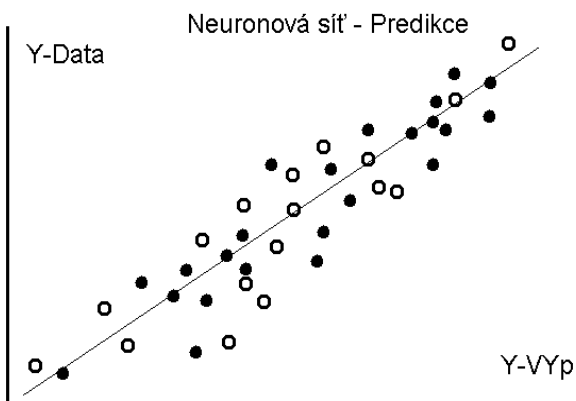


Obrázek 9 Uspokojivá predikční schopnost

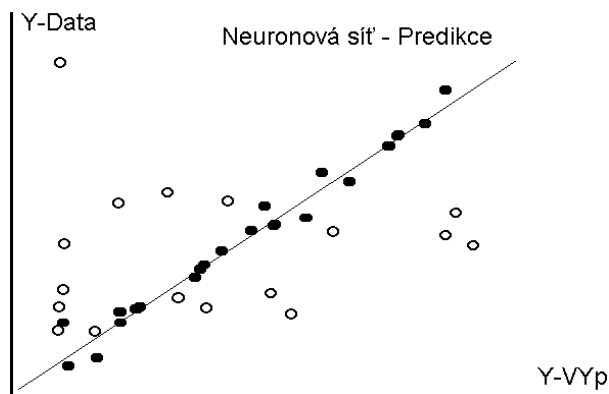


Obrázek 10 Špatná, nebo žádná predikční schopnost, model bude asi nepoužitelný pro predikci – nutno zjednodušit model, nebo přidat data

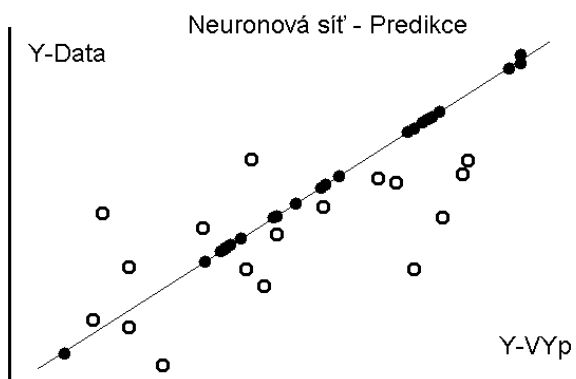
K posouzení kvality predikčních schopností neuronové sítě slouží validace (cross-validation). Validace je založená na jednoduchém principu použít k učení NN určitou část P dat. Tato trénovací, nebo učící část se volí typicky kolem $P = 0.7$, tedy 70%, tento podíl není ovšem nijak závazný. Ze zbývajících částí $1-P$ dat (testovací, či validační data) se pak po, nebo i během optimalizace počítá predikce, která se srovná se skutečnou hodnotou odezvy. Pokud se tato predikce shoduje se skutečnou odezvou, svědčí to o schopnosti sítě predikovat správně odezvu i pro data, která předtím “neviděla”. Kvalitu predikce lze posoudit pomocí grafu průběhu optimalizace nebo grafu predikce. Následující Obrázek 8 až Obrázek 10 ilustruje tři modelové situace s použitím 30% náhodně vybraných validačních dat, tedy $P = 0.7$. Obrázek 11 až Obrázek 14, str. 5 pak ilustruje různé predikční schopnosti NN-modelů na grafu Data-Predikce.



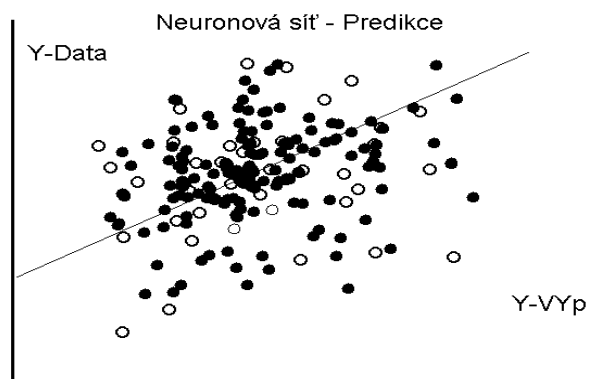
Obrázek 11 Srovnatelná kvalita predikce pro trénovací i testovací data. Dobrá predikční schopnost



Obrázek 12 Dobrá predikce pro trénovací data, špatná pro testovací data - špatná predikční schopnost



Obrázek 13 Dokonalý fit pro trénovací data, velmi špatná predikce, zřejmě přeurčený model, nutno zjednodušit síť



Obrázek 14 Nevýrazná závislost, buď prostě neexistuje závislost odezvy na prediktoru, nebo příliš jednoduchý model, nebo se nepodařilo nalézt parametry sítě.

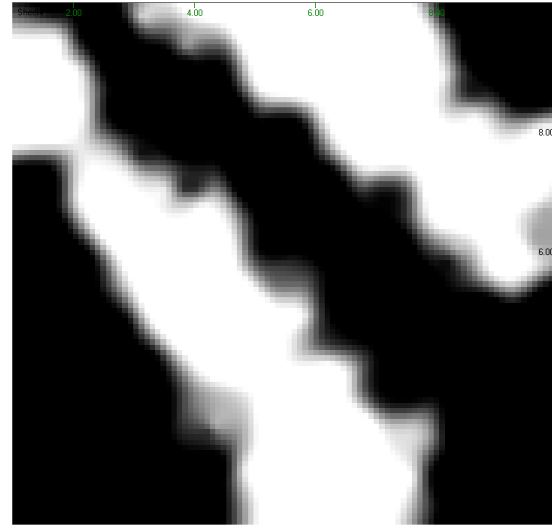
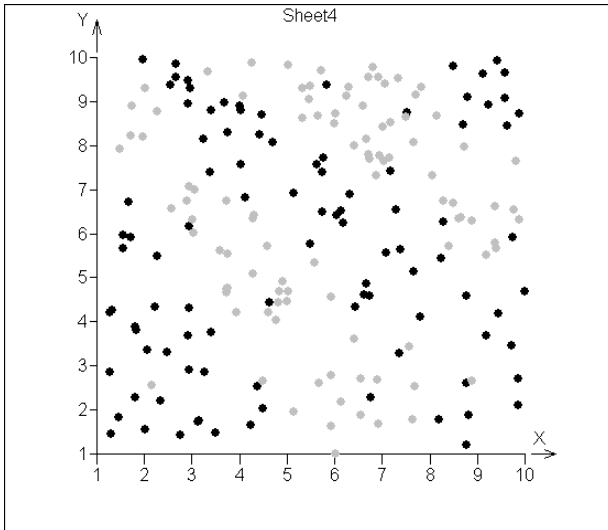
Klasifikační úlohy

Vzhledem k přechodové povaze logistické aktivační funkce jsou neuronové sítě vhodné mimo jiné i k modelování klasifikačních úloh, kdy má výstupní proměnná diskretní hodnotu, nejčastěji dvouúrovňovou (binární), 0 a 1, případně víceúrovňovou, jako např. 1, 2, 3, jejíž výskyt závisí na nezávisle proměnných. Neuronová síť pak predikuje úroveň výstupní proměnné za daných hodnot nezávisle proměnných, podobně jako logistická regrese. V případě binární odezvy 0 – 1 lze predikci považovat za pravděpodobnost nastání stavu “1”. Následující grafy ilustrují použití NN jako klasifikačního modelu. Vlevo jsou naměřené odezvy (světlý bod odpovídá hodnotě 0, tmavý bod hodnotě 1). Vpravo je mapa predikovaných oblastí získaná pomocí neuronové sítě s predikcí zobrazená modulem *Grafy – 3D-Spline*.

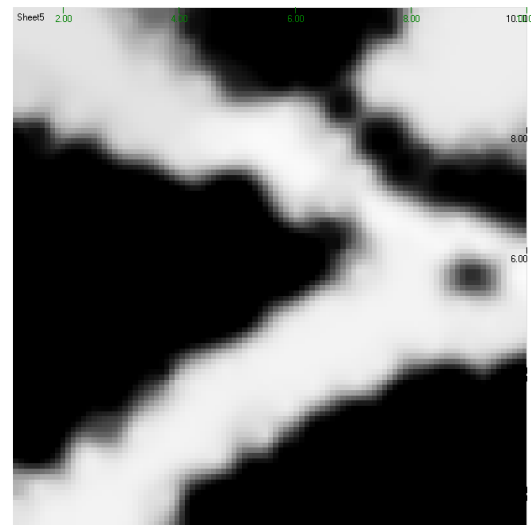
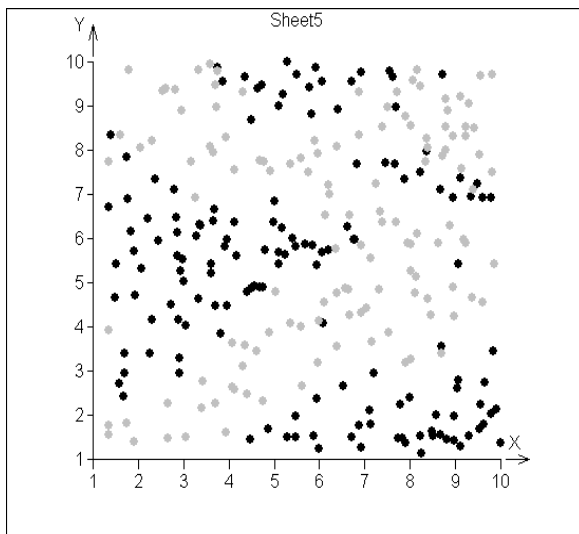
tabulka 1 Příklad dat pro klasifikační úlohu: Parametry X a Y (nezávisle proměnné) ovlivňují výsledek (závisle proměnná, odezva). Při výpočtu je nutno zadat číselné, například binární vyjádření výsledku. Úrovní odezvy může být i více než dvě.

Parametr X	Parametr Y	Výsledek	Výsledek-binární
1.5	3.1	Vyhovuje	0
1.9	2.2	Vyhovuje	0

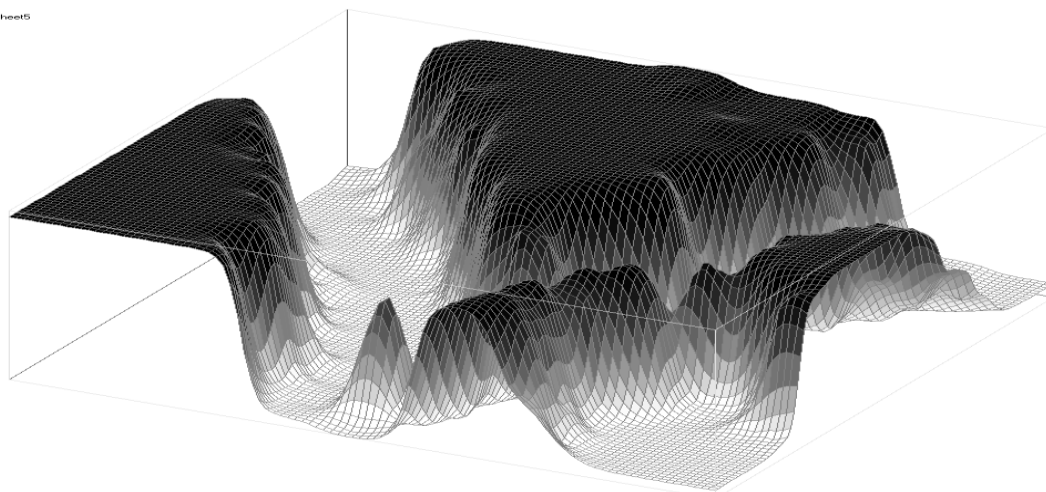
3.5	2.8	Vadný	1
2.9	4.3	Vadný	1
2.4	2.7	Vyhovuje	0
.....



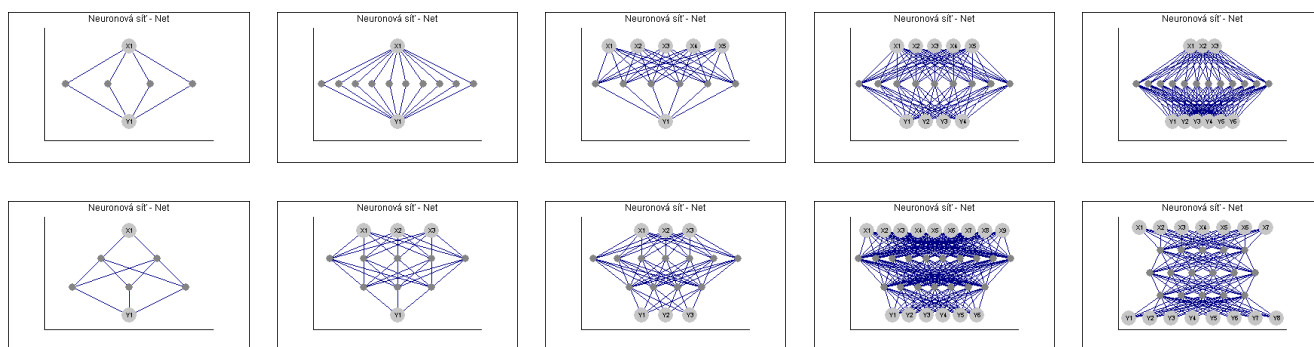
Obrázek 15 Lineárně separabilní data - klasifikační model získaný NN s 1 skrytou vrstvou s 6 neurony



Obrázek 16 Lineárně neseparabilní data - klasifikační model získaný NN s 2 skrytými vrstvou s 5+5 neurony



Obrázek 17 3D reprezentace předchozího grafu klasifikačního modelu pomocí predikce a 3D-Spline (modul Grafy)



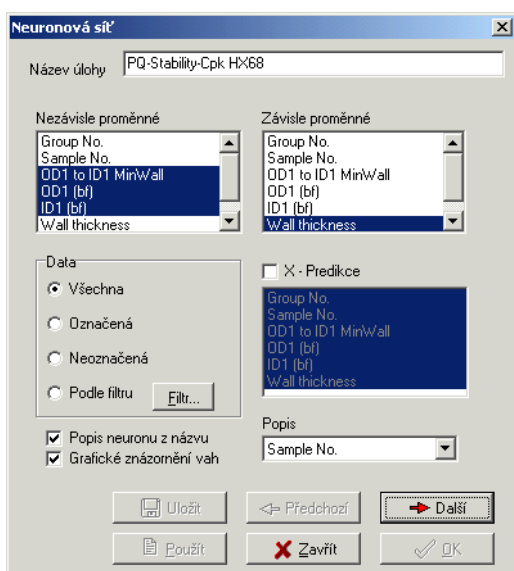
Obrázek 18 Ilustrativní příklady vhodných architektur sítí

Data a parametry

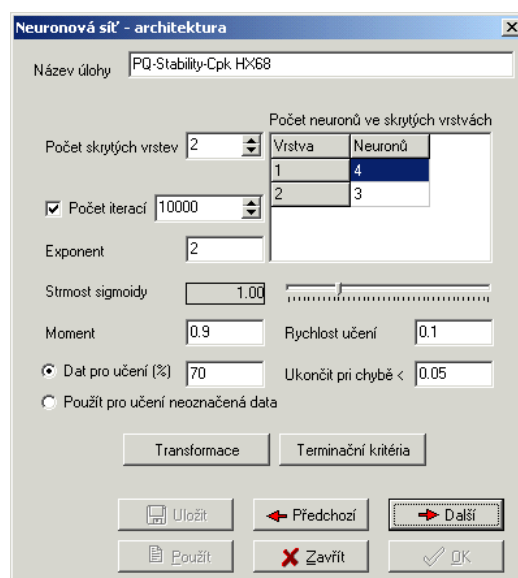
Modul *Neuronové sítě* má několik po sobě jdoucích dialogových oken, v nichž lze nastavit parametry výpočtu. V prvním okně, Obrázek 19, se vyberou sloupce nezávisle a závisle proměnných. V položce *Data* se zvolí, chceme-li k výpočtu použít všechna data, nebo jen nějakou pomnožinu. Zaškrtneme-li políčko *Predikce*, můžeme v tomto poli vybrat nezávisle proměnné, pro které chceme vypočítat predikované hodnoty závisle proměnné. Vybrat musíme stejný počet sloupců, jako jsme vybrali v poli *Nezávisle proměnné*, hodnoty musí mít rovněž podobné hodnoty jako příslušné nezávisle proměnné, aby měla predikce smysl. Vybrat můžeme i tytéž sloupce, jako v poli *Nezávisle proměnné*. Zaškrtneme-li pole *Popis neuronu z názvu*, použije se název sloupce k popisu vstupních a výstupních neuronů v grafu. Je-li zaškrtnuto políčko *Grafické znázornění vah* (doporučeno), znázorní se absolutní hodnota vah jako tloušťka spojovací úsečky mezi neurony (“synapsí”) a znaménko jako barva (modrá = kladná váha, červená = záporná váha). Tlačítkem *Další* se dostaneme do dalšího okna. V okně *Architektura neuronové sítě* se především definuje počet skrytých vrstev, typický počet je 1, 2, nejvýše 3. Více než 3 vrstvy jsou účelné jen v některých speciálních případech. V poli *Počet neuronů ve skrytých vrstvách* určíme, kolik neuronů mají obsahovat jednotlivé vrstvy. *Počet iterací* omezuje délku výpočtu počtem iteračních optimalizačních kroků. Hodnota *Exponent* určuje exponent k kritériální funkce

$$s(NN) = \sum |y - y_{\text{pred}}|^k$$

Doporučená hodnota je 2, hodnoty $1 < k < 2$ mohou poskytnout robustnější modely, které jsou méně ovlivněné případnými hrubými chybami, vybočujícími hodnotami v závisle proměnných. Parametr *Strmost sigmoidy* udává rychlost (citlivost) s jakou budou neurony reagovat na změnu nezávisle proměnných. Doporučená hodnota je 1. Parametry *Moment* a *Rychlost učení* ovlivňují samotný optimalizační algoritmus. Doporučené hodnoty jsou 0.9 a 0.1. V políčku *Dat pro učení (%)* se určí jaká část *P* dat se má použít pro učení sítě. Chceme-li použít cross-validaci, volíme obvykle podíl mezi 50 a 90%. Volba 100% vylučuje cross-validaci a pro výpočet se použijí všechna data. Výběr zvoleného podílu dat se provede generátorem náhodných čísel, vybraných 100.*P* % dat je tedy při každém výpočtu jiných. Pokud chceme použít pevnou podmnožinu pro učení a validaci sítě, můžeme použít volbu *Použít pro učení neoznačená data*. Tlačítkem *Uložit model* se aktuálně vypočítaný model uloží do souboru pro pozdější použití v modulu *Predikce*, případně k automatickým predikcím v inteligentní databázi QCE-DataCenter®.

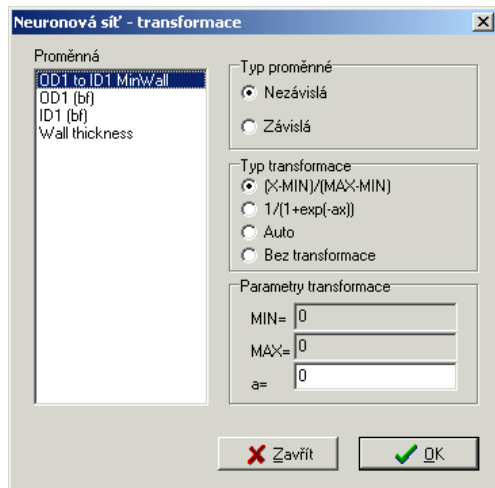


Obrázek 19 Neuronová síť - okno pro zadání dat

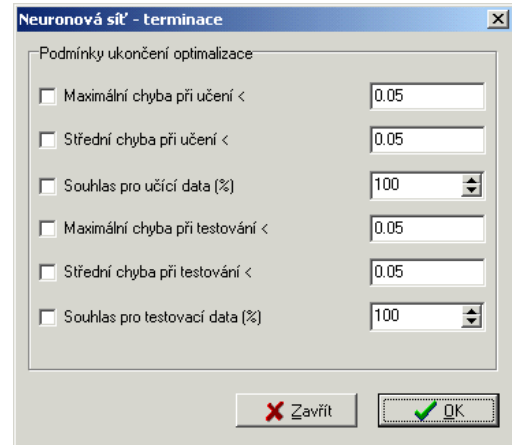


Obrázek 20 Neuronová síť - okno definice architektury sítě a parametrů výpočtu

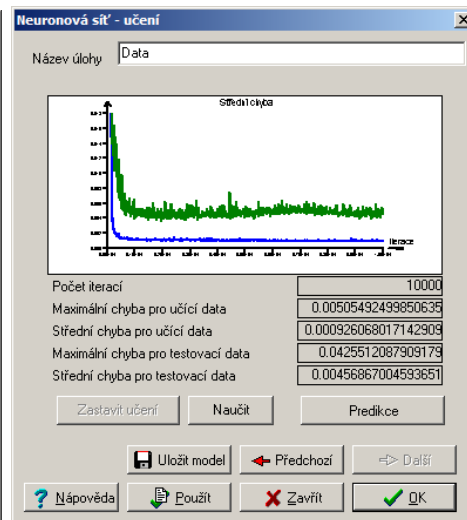
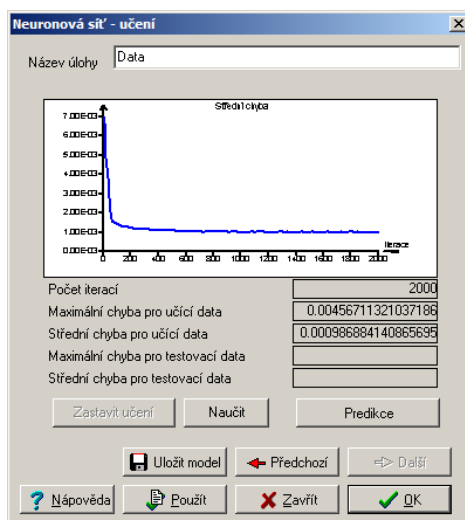
Další postup se řídí následujícími okny.



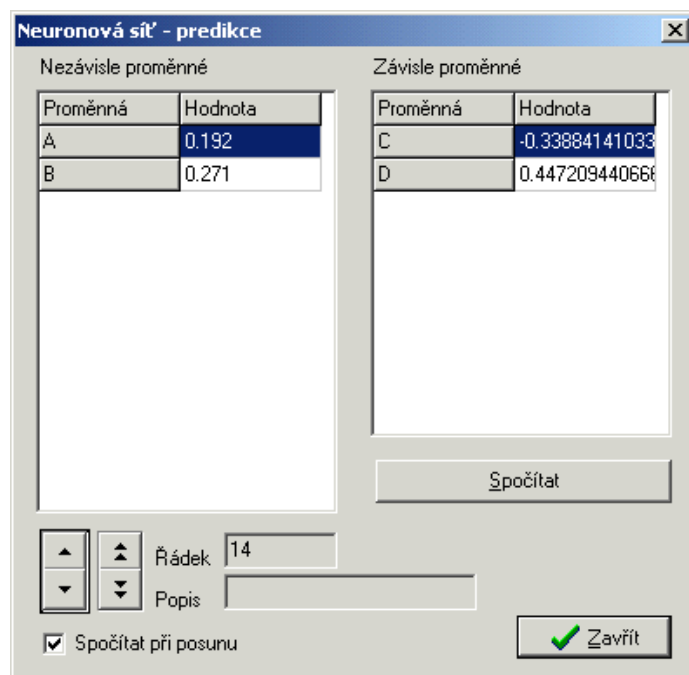
Obrázek 21 Transformace vstupních proměnných (prediktorů)



Obrázek 22 Terminační kritéria pro optimalizaci neuronové sítě



Obrázek 23 Neuronová síť - průběh učení sítě bez validace a s validací na testovacích datech

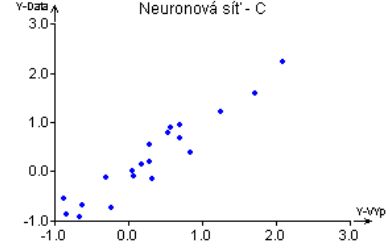
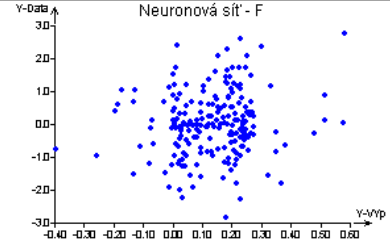
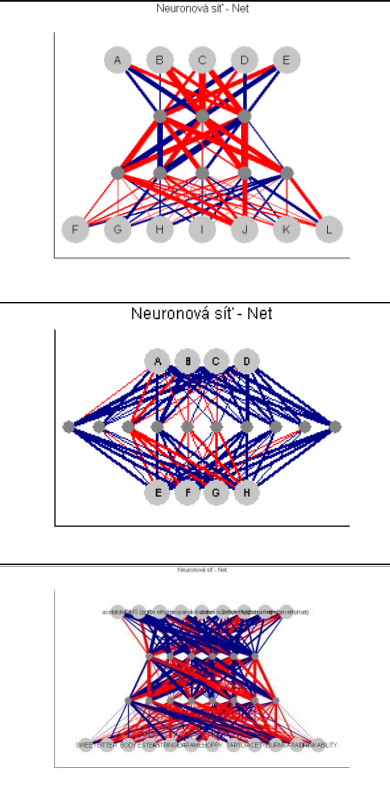
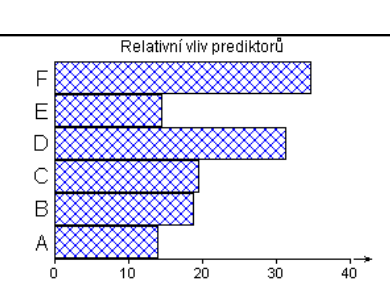


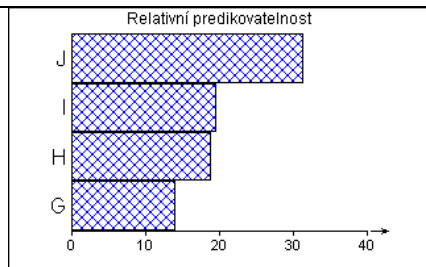
Obrázek 24 Predikce pomocí neuronové sítě

Protokol

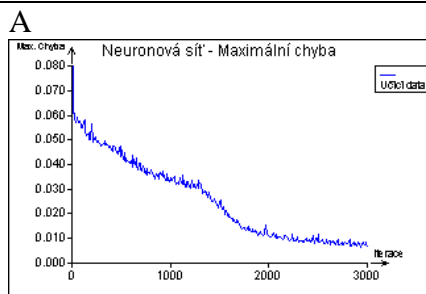
Název úlohy	Zadaný název úlohy
Data	
Nezávisle proměnné	Seznam nezávisle proměnných
Typ transformace	Typ transformace nezávisle proměnných
Závisle proměnné	Seznam závisle proměnných
Typ transformace	Typ transformace závisle proměnných
Vrstva, Neuronů	Číslo vrstvy a počet neuronů ve vrstvě
Strmost sigmoidy	Zadaná strmost sigmoidy
Moment	Zadaný moment
Rychlost učení	Zadaná rychlost učení
Ukončit při chybě	Kritérium chyby pro ukončení
Procent dat pro učení (%)	Procent dat pro učení (%), je-li vybráno
Podmínky ukončení optimalizace	Zadané terminační podmínky pro ukončení výpočtu
Výpočet	Informace o průběhu výpočtu
Počet iterací	Skutečný počet iterací
Maximální chyba pro učící data	Dosažená maximální chyba pro učící data
Střední chyba pro učící data	Dosažená střední chyba pro učící data
Maximální chyba pro testovací data	Dosažená maximální chyba pro testovací data
Střední chyba pro testovací data	Dosažená střední chyba pro testovací data
Váhy	Vypočítané optimální váhy neuronové sítě
Vrstva / Neuron	Číslo vrstvy a číslo neuronu
Predikce	Tabulka predikovaných hodnot, byla-li zadána predikce.

Grafy

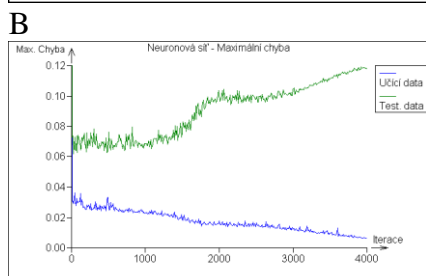
	<p>Graf Y-Predikce. Graf shody naměřených (zadaných) hodnot závisle proměnné a vypočítaných hodnot predikce. Čím více se body blíží přímce, tím lepší je fit, tedy predikce dobře odpovídá skutečným hodnotám. Tento graf se vytváří pro každou závisle proměnnou. Některé proměnné může NN predikovat lépe, jiné hůře.</p>
	<p>Nemá-li graf tvar zřetelného trendu, nebyl nalezen model, který by dobře predikoval tuto nezávisle proměnnou. Příčinou může být buď předčasné ukončení výpočtu před dosažením optimálních vah, nebo příliš jednoduchá síť, která není schopna postihnout případné složitější závislosti, nebo tato proměnná na zvolených prediktorech výrazně nezávisí.</p>
	<p>Grafické vyjádření architektury sítě. Byla-li při výpočtu vybrána možnost <i>Grafické znázornění vah</i>, pak tloušťka synapsí znázorňuje absolutní velikost odpovídající váhy a tím také množství informace, která teče mezi dvěma neurony. Podle tloušťky synapse se tak dá usuzovat na významnost nezávisle proměnné (vychází z ní tlusté synapse) a také na kvalitu predikce jednotlivých závisle proměnných (vchází do ní tlusté synapse). Barva synapsí určuje pouze znaménko váhy (červená = záporná váha, modrá = kladná váha). Byla-li vybrána možnost <i>Popis neuronu z názvu</i>, použije se název sloupce jako popis vstupních a výstupních neuronů.</p>
	<p>Relativní vliv prediktorů na predikci vyjádřený jako součet absolutních vah jednotlivých prediktorových proměnných. Tato statistika je pouze mírou vlivu proměnných na odezvu, není testem významnosti. Tento graf má reálnou vypovídací schopnost pouze při vhodné volbě modelu.</p>



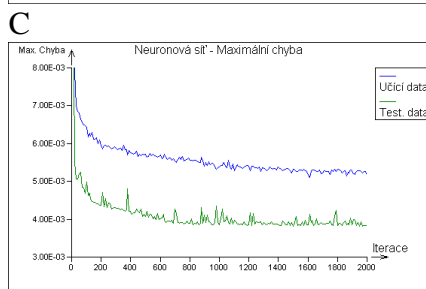
Relativní predikovatelnost je součet absolutních vah jednotlivých výstupních proměnných a vyjadřuje míru ovlivnění těchto proměnných zvolenými prediktory. Tato statistika není potvrzením statistické významnosti vztahu mezi predikcí a odezvou. Tento graf má reálnou vypovídací schopnost pouze při vhodné volbě modelu.



Graf učícího procesu, pokles součtu čtverců rozdílů predikce a skutečných hodnot závisle proměnné v závislosti na počtu iterací. Typické průběhy jsou podrobněji popsány v předchozích odstavcích. Pokud byla zvolena cross-validace, tedy v poli *Dat pro učení*, Obrázek 20, kreslí se také chyba predikce. Podle vývoje maximální chyby predikce se dá ještě v průběhu optimalizace (učení) usuzovat o kvalitě modelu a dat.



Na obrázku **A** je typický úspěšný učící proces, který postupně zlepšoval model pro zadaná data, ovšem bez validace, tedy bez hodnocení predikční schopnosti modelu pro nová data. Na obrázku **B** se sice rovněž daří zlepšovat model (klesající chyba, spodní křivka). Ale predikce pro neznámá data se zhoršuje (rostoucí chyba predikce, horní křivka) Taková situace obvykle souvisí s malým počtem dat, nebo s přeuročností (nadbytečnou složitostí) použitého modelu.



Křivka **C** svědčí o dobré kvalitě modelu, který bude dobře predikovat z neznámých dat, obě křivky ukazují zlepšování jak modelu, tak i predikce validačních dat. Křivka **D** nevykazuje v průběhu učení žádné zlepšení. To může být způsobeno neexistencí závislosti v datech, nebo příliš jednoduchým, nebo i příliš složitým modelem, který se nepodařilo optimalizovat.

