

Systemy pomiarowo-diagnostyczne

Metody uczenia maszynowego – wykład III

bogumil.konopka@pwr.edu.pl

2016/2017

Wykład III - plan

- Regresja logistyczna
- Ocena skuteczności klasyfikacji
 - Macierze pomyłek
 - Krzywe ROC
- Sieci neuronowe
 - Propagacja informacji
 - Algorytm gradientu prostego
 - Wsteczna propagacja błędu

Klasyfikacja, a model liniowy

- Problemy z modelem liniowym:
 - W modelu liniowym $\hat{f}(x)$ może być >1 lub < 0
 - Interpretacja jest trudna
 - Problemem jest również określenie parametru odcięcia
- Przy klasyfikacji chcemy żeby:
$$0 \leq \hat{f}(x) \leq 1$$



Regresja logistyczna

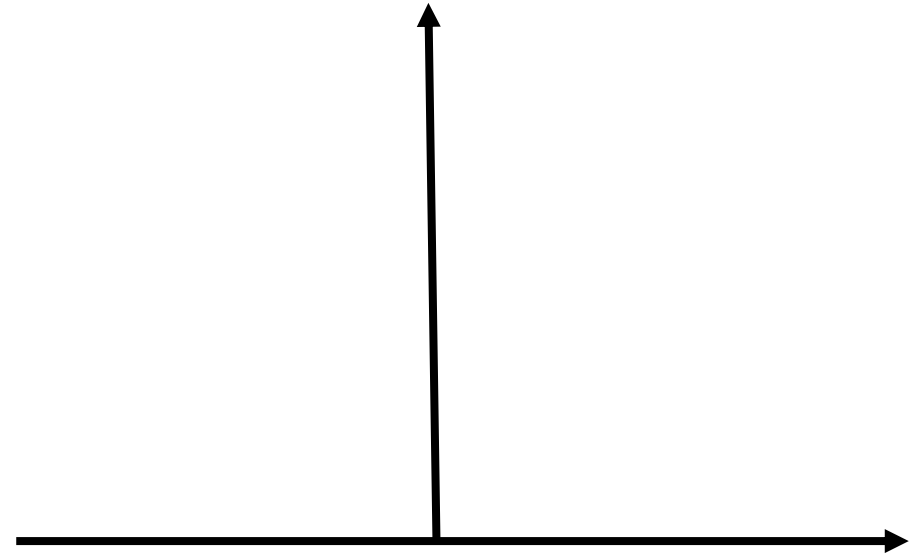
- Przy klasyfikacji chcemy żeby:

$$0 \leq \hat{f}(x) \leq 1$$

- Funkcja logistyczna:

$$\hat{f}(x) = \frac{1}{1 + e^{-z}} \leftarrow$$

$$\hat{f}(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}}$$



Interpretacja modelu logistycznego

- $\hat{f}(x) = P(y = 1|X)$ - prawdopodobieństwo, tego że $y = 1$ przy zadanych wartościach x

- Przykład:

Jeżeli, dla modelu klasyfikującego nowotwory

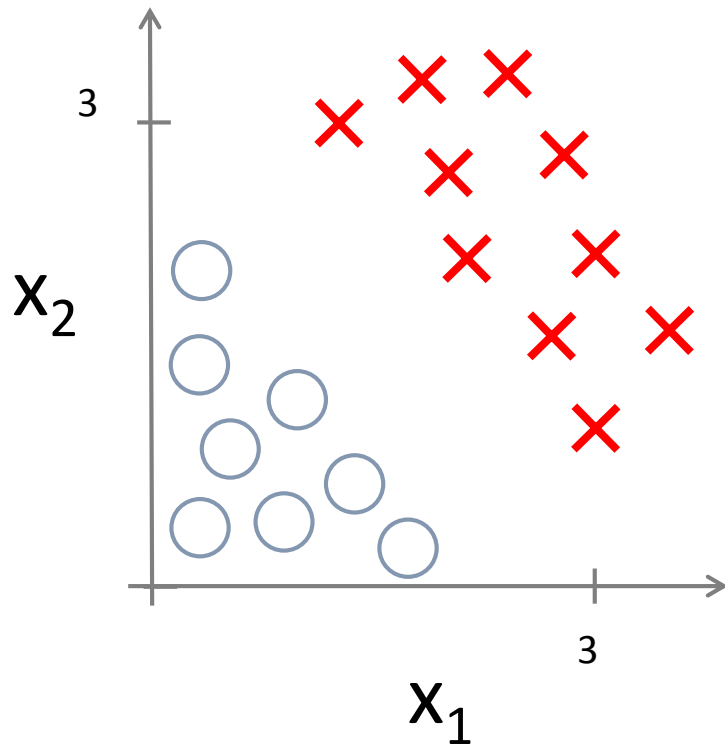
$$\hat{f}(x) = 0.7$$

To możemy powiedzieć:

„Jest 70% szans na to, że nowotwór jest złośliwy”

Graficzna interpretacja ,z'

Dane w przestrzeni atrybutów

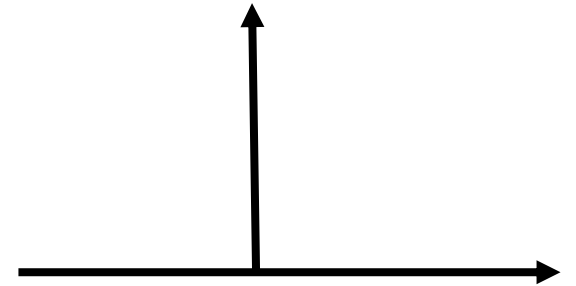


$$\hat{f}(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2)}}$$

Model dopasowany do danych:

$$\hat{f}(x) = \frac{1}{1 + e^{-(-3 + x_1 + x_2)}}$$

Kiedy $y = 1$?



Granica decyzji:

Dopasowanie współczynników

- Najczęściej współczynniki modelu dopasowuje się *metodą maksymalnej wiarygodności* (maximum likelihood)
- Intuicja algorytmu:
 - Wyszukiwane są takie wartości współczynników, dla których prawdopodobieństwo uzyskania danego przyporządkowania do klas jest największe
 - Maksymalizowana funkcja ma postać:

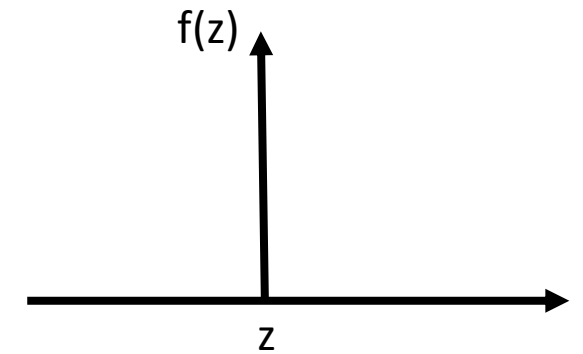
$$l(\beta) = \prod_{i:y_i=1} p(x_i; \beta) \prod_{j:y_j=0} (1 - p(x_j; \beta))$$

Ocena dokładności współczynników

- Współczynniki ocenia tak jak dla regresji liniowej:
 - Przedziały ufności
 - Test t-Studenta

	Coefficient	Std. error	Z-statistic	P-value
Intercept	-10.8690	0.4923	-22.08	<0.0001
balance	0.0057	0.0002	24.74	<0.0001
income	0.0030	0.0082	0.37	0.7115
student[Yes]	-0.6468	0.2362	-2.74	0.0062

TABLE 4.3. For the **Default** data, estimated coefficients of the logistic regression model that predicts the probability of **default** using **balance**, **income**, and **student** status. Student status is encoded as a dummy variable **student[Yes]**, with a value of 1 for a student and a value of 0 for a non-student. In fitting this model, **income** was measured in thousands of dollars.



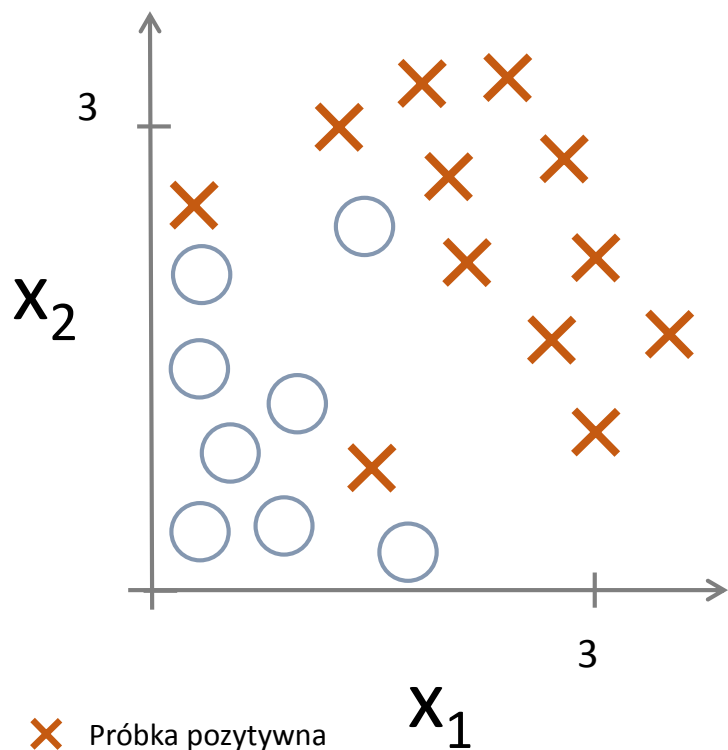
$$\hat{f}(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2)}}$$

Zadanie domowe na punkt z aktywności

- Jaka jest interpretacja współczynników w modelu regresji logistycznej?
- Wskazówki:
 - β_0 :Wychodząc od $\hat{f}(x) = \frac{1}{1+e^{-(\beta_0+\beta_1x_1)}}$ wyprowadzić β_0 lub e^{β_0} , pamiętając, że $\hat{f}(x) = p(y = 1|x)$
 - β_1 : Wyprowadzić e^{β_1} wyliczając stosunek $\frac{e^{-(\beta_0+\beta_1x_1)}}{e^{-(\beta_0+\beta_1x'_1)}}$, gdy $x'_1 = x_1 + 1$

Ocena dokładności klasyfikatora (klasyfikator binarny)

Dane w przestrzeni atrybutów

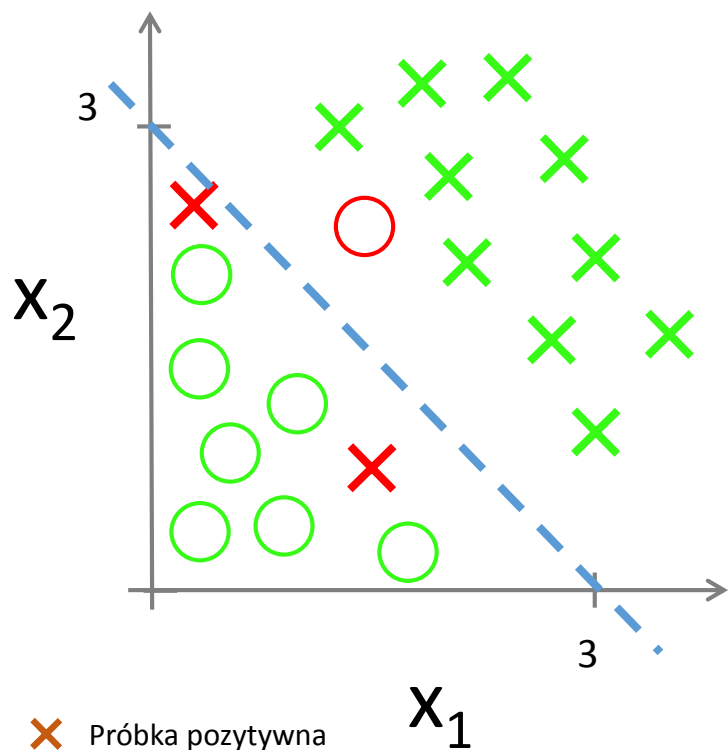


- **Możliwe wyniki klasyfikacji:**

- **TP** – **True Positive** – Prawidłowo zaklasyfikowany przypadek pozytywny
- **FP** – **False Positive** – Fałszywy pozytywny – Przypadek negatywny zaklasyfikowany jako pozytywny
- **TN** – **True Negative** – Prawidłowo zaklasyfikowany przypadek negatywny
- **FN** – **False Negative** – Fałszywy negatywny – Przypadek pozytywny zaklasyfikowany jako negatywny

Ocena dokładności klasyfikatora (klasyfikator binarny)

Dane w przestrzeni atrybutów



- **Możliwe wyniki klasyfikacji:**

- **TP** – **True Positive** – Prawidłowo zaklasyfikowany przypadek pozytywny
- **FP** – **False Positive** – Fałszywy pozytywny – Przypadek negatywny zaklasyfikowany jako pozytywny
- **TN** – **True Negative** – Prawidłowo zaklasyfikowany przypadek negatywny
- **FN** – **False Negative** – Fałszywy negatywny – Przypadek pozytywny zaklasyfikowany jako negatywny

Macierz pomyłek

		<u>Rzeczywista klasa</u>		
		Pozytywna	Negatywna	
<u>Przewidywana klasa</u>	Pozytywna	TP	FP	$\frac{TP}{P_{predicted}}$
	Negatywna	FN	TN	$\frac{TN}{N_{predicted}}$
		$\frac{TP}{P_{total}}$	$\frac{TN}{N_{total}}$	$\frac{TP + TN}{P_{total} + N_{total}}$

Sensitivity - czułość

Specificity - swoistość

Positive Predictive Value/Precision – celność przewidywania przypadków pozytywnych

Negative Predictive Value – celność przewidywania przypadków negatywnych

Accuracy (ACC) – dokładność

Macierz pomyłek

		Rzeczywista klasa		
		Pozytywna	Negatywna	
Przewidywana klasa	Pozytywna	TP	FP	$\frac{TP}{P_{predicted}}$
	Negatywna	FN	TN	$\frac{TN}{N_{predicted}}$
		$\frac{TP}{P_{total}}$	$\frac{TN}{N_{total}}$	$\frac{TP + TN}{P_{total} + N_{total}}$

$$\frac{FP}{N_{total}}$$

False Positive Rate –
odsetek predykcji
fałszywie pozytywnych

Positive Predictive Value/Precision –
celność przewidywania przypadków
pozytywnych

Negative Predictive Value –
celność przewidywania
przypadków negatywnych

Accuracy (ACC) – dokładność

Sensitivity - czułość

Specificity - swoistość

Pełna macierz pomyłek z wszystkimi parametrami

		Condition (as determined by "Gold standard")			
		Condition positive	Condition negative		
Total population				Prevalence = $\frac{\sum \text{Condition positive}}{\sum \text{Total population}}$	
Test outcome	Test outcome positive	True positive	False positive (Type I error)	Positive predictive value (PPV), Precision = $\frac{\sum \text{True positive}}{\sum \text{Test outcome positive}}$	False discovery rate (FDR) = $\frac{\sum \text{False positive}}{\sum \text{Test outcome positive}}$
	Test outcome negative	False negative (Type II error)	True negative	False omission rate (FOR) = $\frac{\sum \text{False negative}}{\sum \text{Test outcome negative}}$	Negative predictive value (NPV) = $\frac{\sum \text{True negative}}{\sum \text{Test outcome negative}}$
Accuracy (ACC) = $\frac{\sum \text{True positive} + \sum \text{True negative}}{\sum \text{Total population}}$		True positive rate (TPR), Sensitivity, Recall = $\frac{\sum \text{True positive}}{\sum \text{Condition positive}}$	False positive rate (FPR), Fall-out = $\frac{\sum \text{False positive}}{\sum \text{Condition negative}}$	Positive likelihood ratio (LR+) = $\frac{\text{TPR}}{\text{FPR}}$	Diagnostic odds ratio (DOR) = $\frac{\text{LR+}}{\text{LR-}}$
		False negative rate (FNR) = $\frac{\sum \text{False negative}}{\sum \text{Condition positive}}$	True negative rate (TNR), Specificity (SPC) = $\frac{\sum \text{True negative}}{\sum \text{Condition negative}}$	Negative likelihood ratio (LR-) = $\frac{\text{FNR}}{\text{TNR}}$	

http://en.wikipedia.org/wiki/Sensitivity_and_specificity

Dlaczego ogólna skuteczność klasyfikacji (ACC) nie wystarcza?

Zbiór liczy **50** pacjentów chorych (P) oraz **50** zdrowych (N)

		Rzeczywista klasa		
		P = 50	N = 50	
Przewidywana klasa	P	45		
	N		47	

Zbiór liczy **1000** pacjentów chorych (P) oraz **100** zdrowych (N)

		Rzeczywista klasa		
		P = 1000	N = 100	
Przewidywana klasa	P	997		
	N		1	

Dlaczego ogólna skuteczność klasyfikacji (ACC) nie wystarcza?

Zbiór liczy **50** pacjentów chorych (P) oraz **50** zdrowych (N)

		Rzeczywista klasa		
		P = 50	N = 50	
Przewidywana klasa	P	45		
	N		47	

Przy równolicznych zbiorach próbek pozytywnych i negatywnych Acc jest ok.

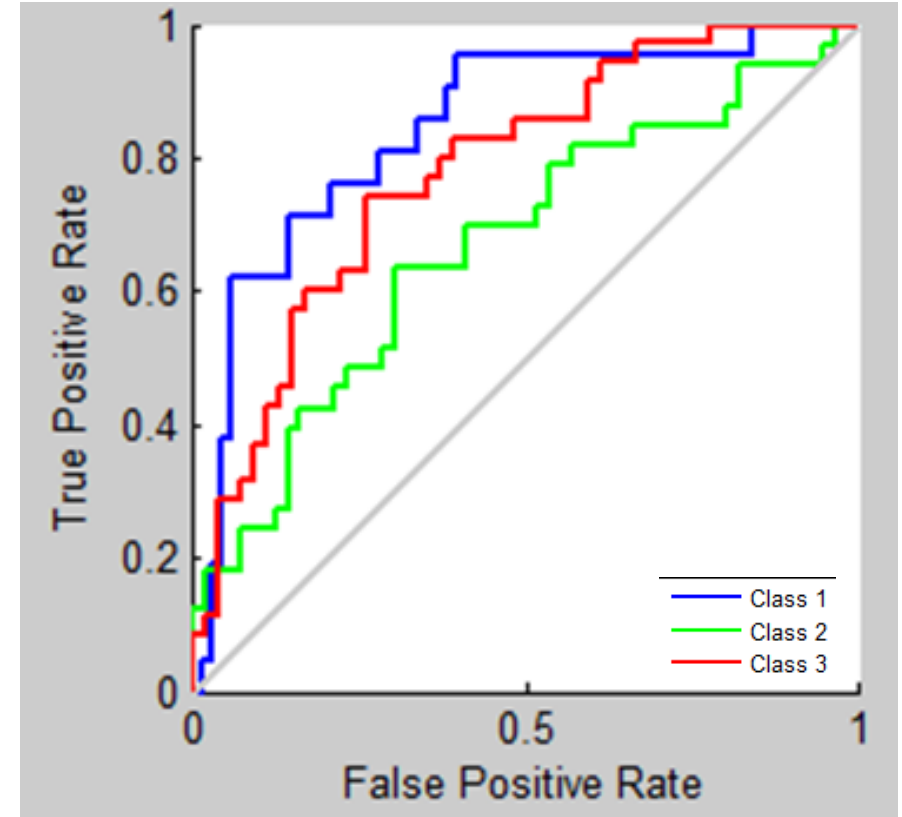
Zbiór liczy **1000** pacjentów chorych (P) oraz **100** zdrowych (N)

		Rzeczywista klasa		
		P = 1000	N = 100	
Przewidywana klasa	P	997		
	N		1	

Jeżeli jedna klasa jest **zdecydowanie** nadreprezentowana, wówczas Acc może przeceniać lub nie doceniać klasyfikator.

Krzywa ROC (Receiver Operating Characteristic)

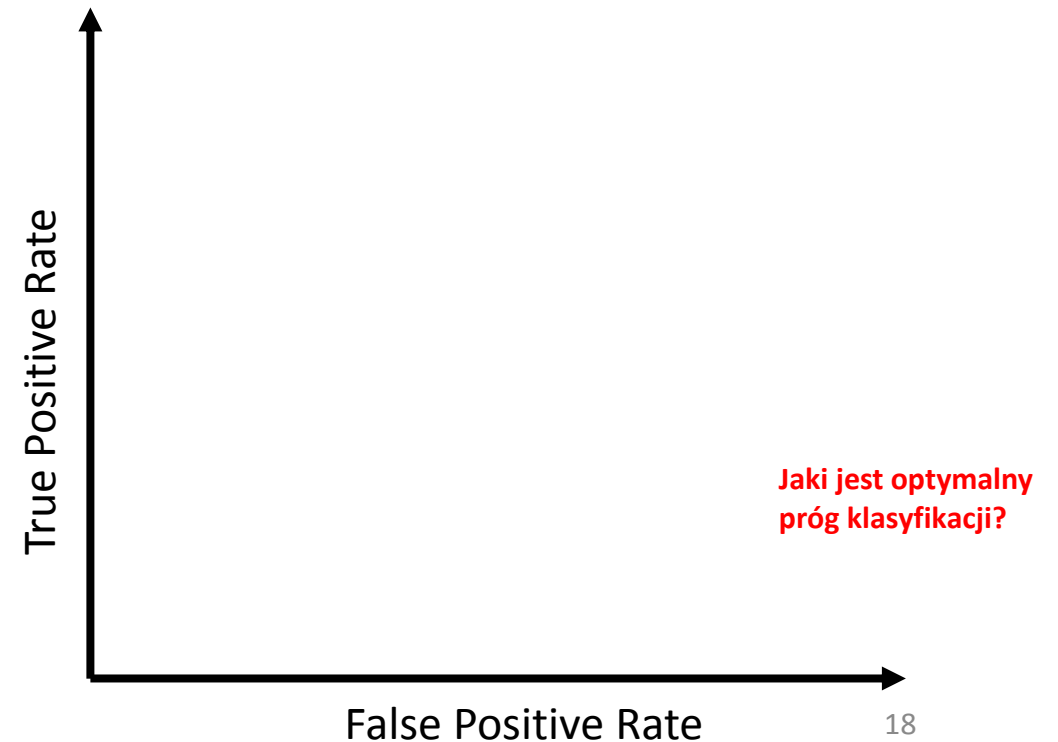
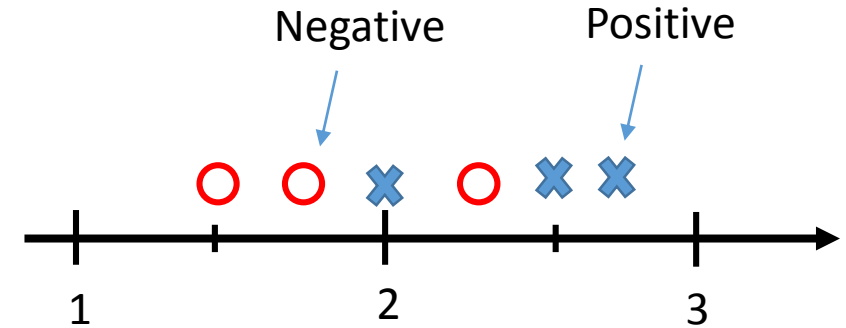
- Cel stosowania:
 - Ocena klasyfikatora
 - Porównywanie klasyfikatorów
 - Wybór optymalnego progu odcięcia
- Pole pod krzywą (Area Under Curve - AUC) mówi o ogólnej skuteczności klasyfikatora
 - 1.0 – 0.9 bardzo dobry
 - 0.9 – 0.8 dobra
 - 0.8 – 0.7 dość dobra
 - 0.7 – 0.6 słaby
 - 0.6 – 0.5 bardzo słaby



Krzywa ROC - algorytm

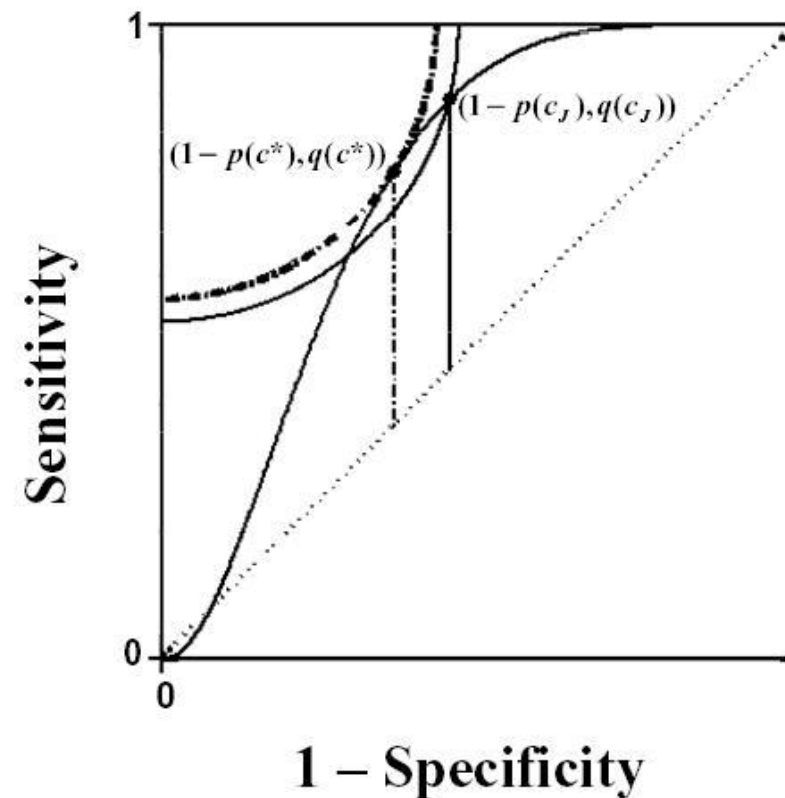
- Uszeregować od największego do najmniejszego
- Obliczać TPR i FPR przy kolejnych wartościach stężenia hormonu jako progach podziału na klasy

$$TPR = \frac{TP}{P_{total}} \quad FPR = \frac{FP}{N_{total}} \quad P_{total} = 3 \times$$
$$N_{total} = 3 \circ$$



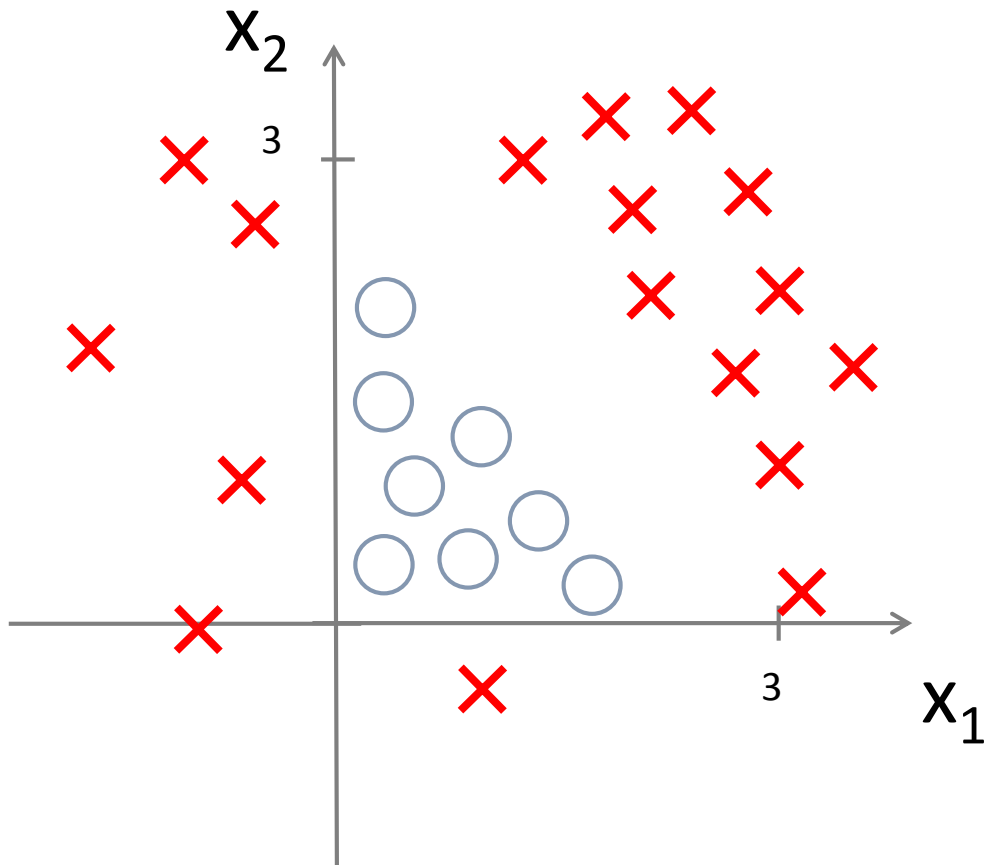
Dobór optymalnego progu klasyfikacji

- Minimalizacja odległości od punktu $(0,1)$ na krzywej ROC
- Maksymalizacja różnicy pomiędzy True Positive Rate i False Positive Rate – tzw. Indeks Youden'a



Klasyfikacja nieliniowa

Dane w przestrzeni atrybutów



Liniowa granica decyzji:

$$\hat{f}(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2)}}$$

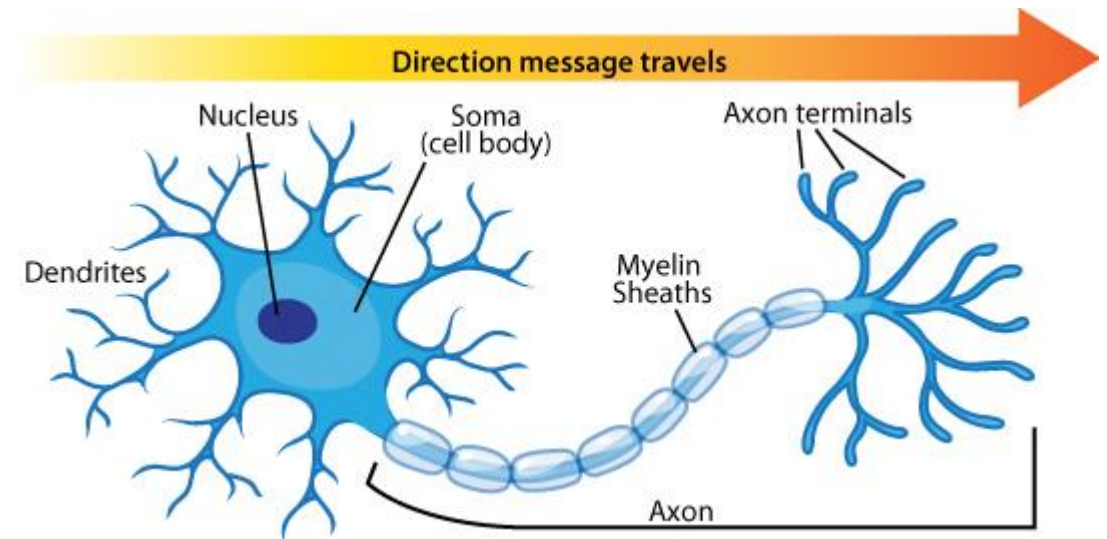
Nieliniowa granica decyzji:

$$\hat{f}(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \beta_4 x_2^2)}}$$

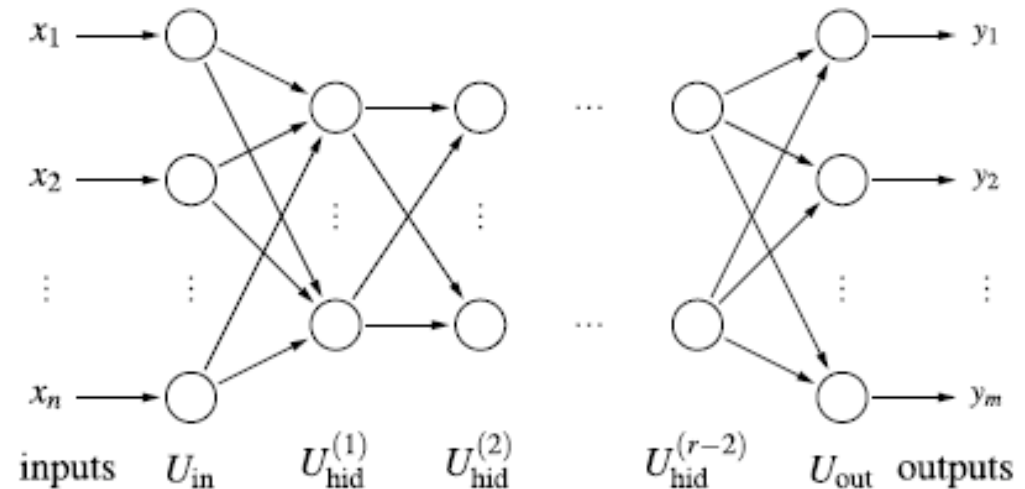
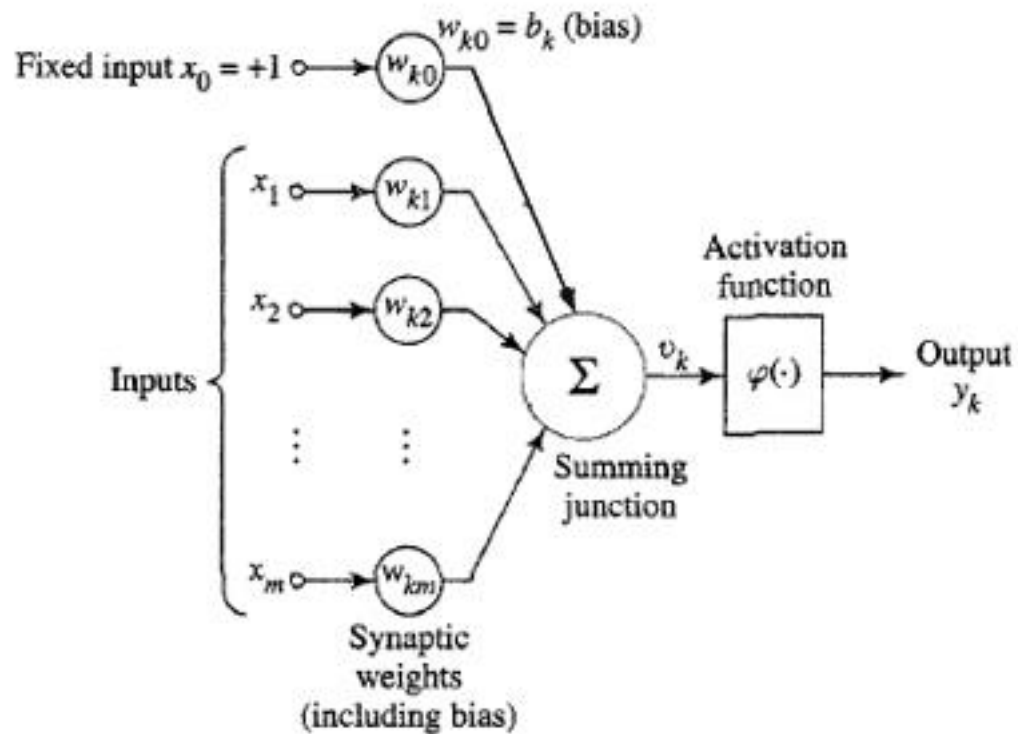
Jak wybrać formułę funkcji?

Sztuczne sieci neuronowe - inspiracja

- Inspiracja biologiczna
 - Neuron otrzymuje bodźce od innych neuronów
 - Bodźce mogą podwyższać lub obniżać poziom potencjału elektrycznego
 - Jeżeli poziom przekroczy określony próg, neuron jest aktywowany i wyzwala impuls
 - Siła wpływu impulsu na inne neurony zależy od warunków chemicznych w synapsie pomiędzy neuronami
 - Modyfikacja połączeń umożliwia sieci uczenie się i rozpoznawanie wzorców pobudzeń

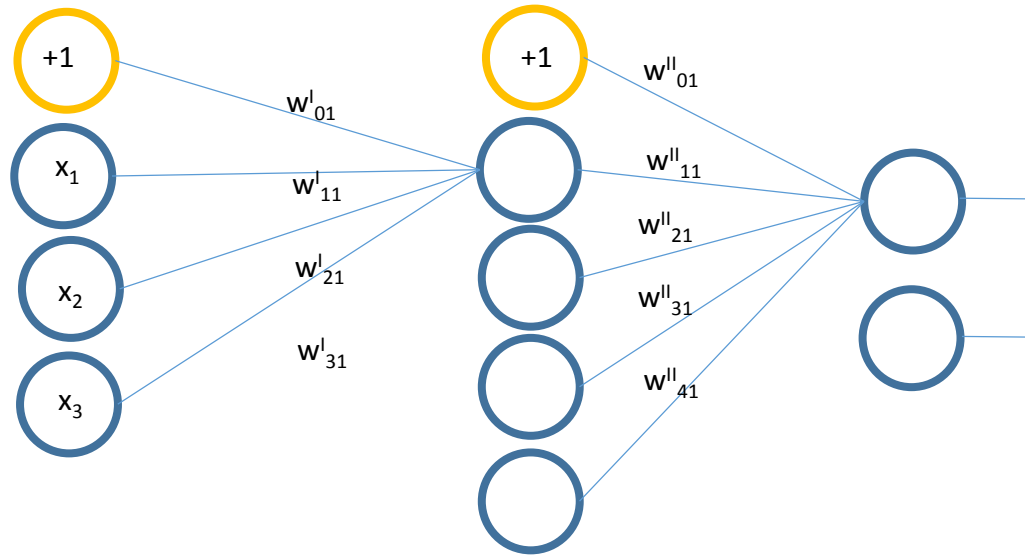


Sztuczna sieć neuronowa - architektura



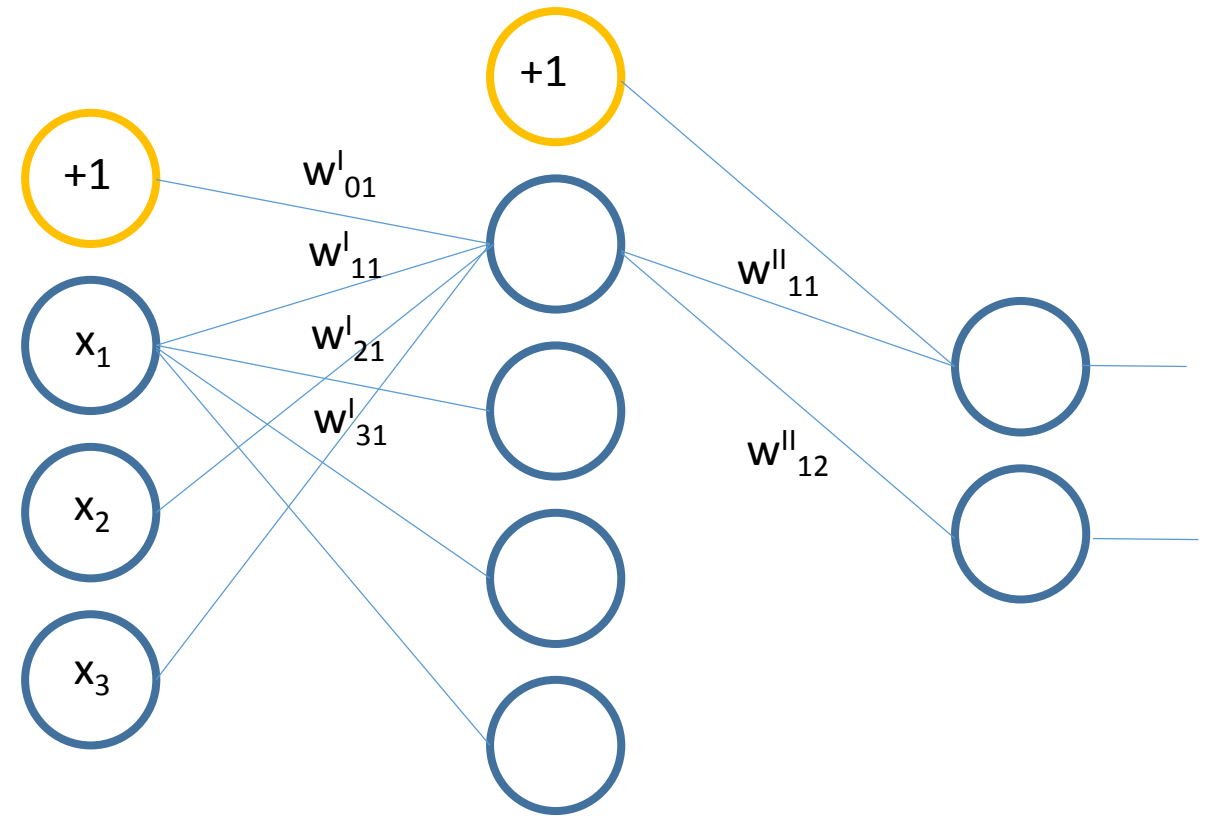
Sztuczna sieć neuronowa - obliczenia

- Obliczenia w sieci „feed-forward”



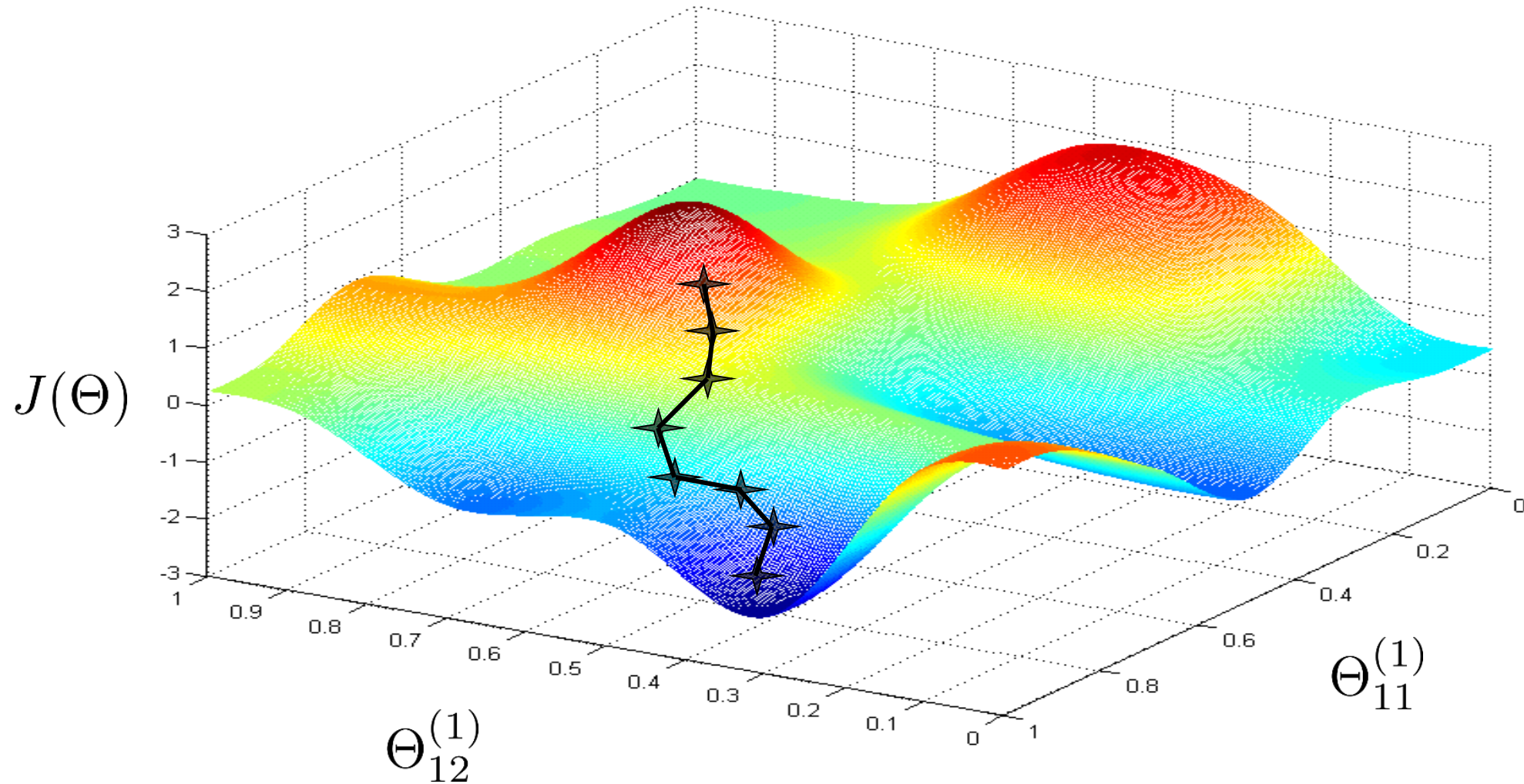
Algorytm gradientu prostego (gradient descent) dla regresji liniowej

Algorytm gradientu prostego dla sieci neuronowych i wsteczna propagacja błędu



(część połączeń na rysunku pominięto dla zachowania czytelności)

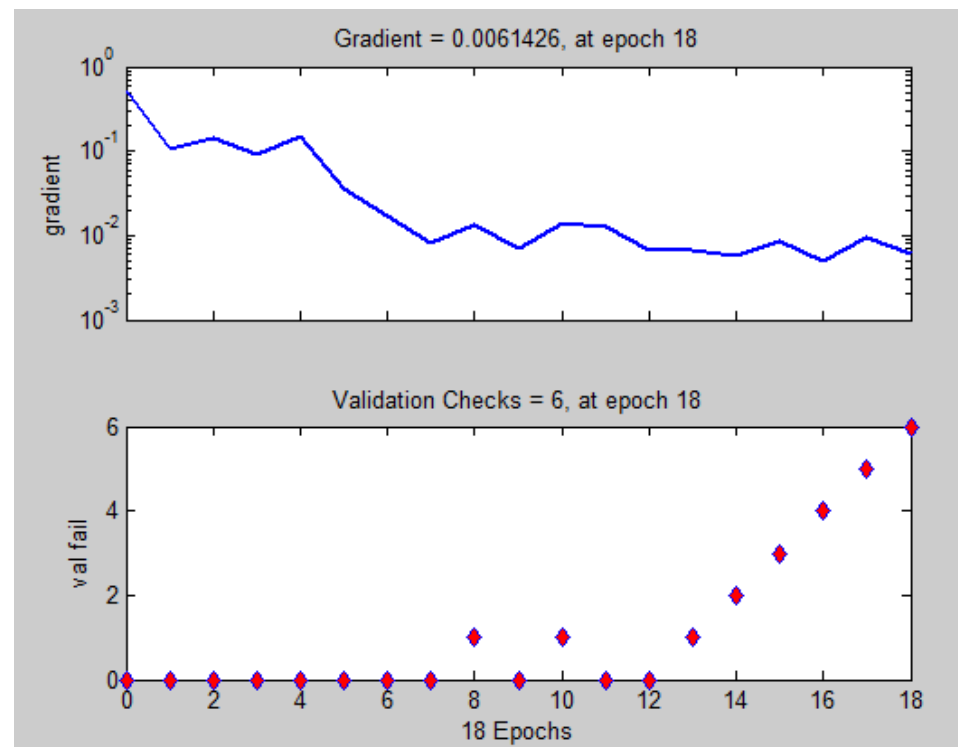
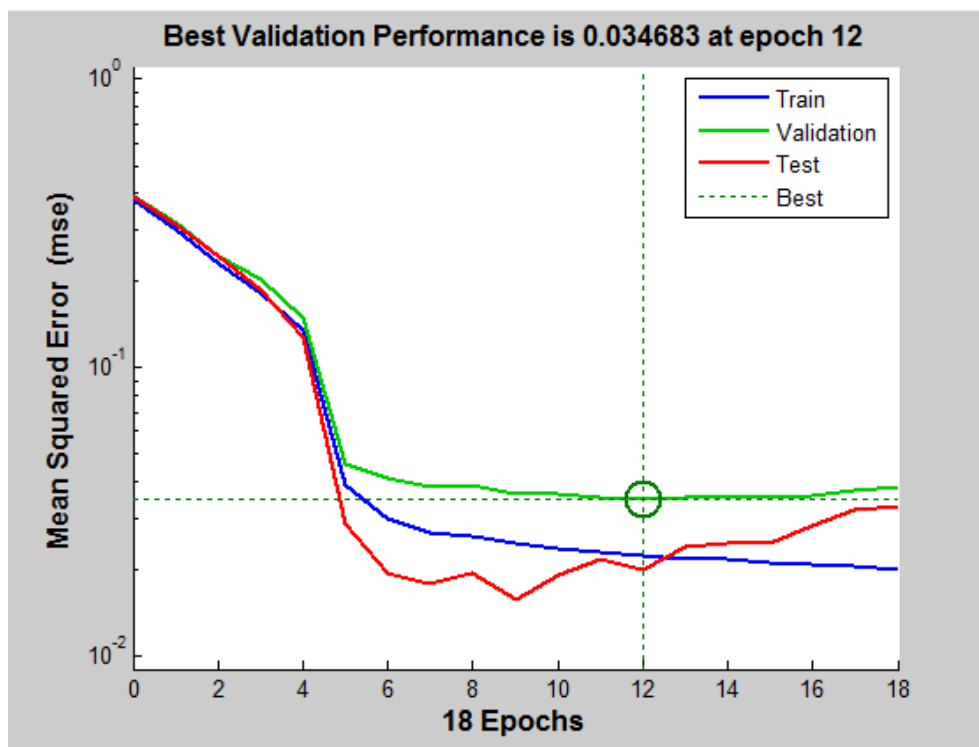
Algorytm gradientu prostego znajduje lokalne minima funkcji



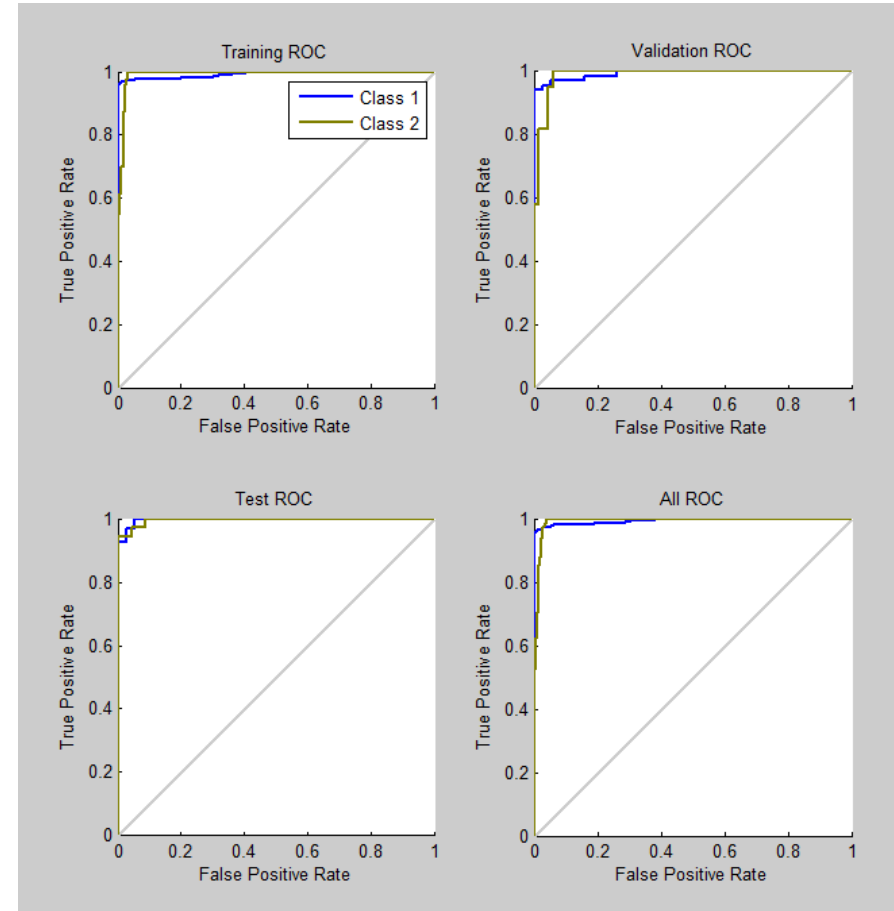
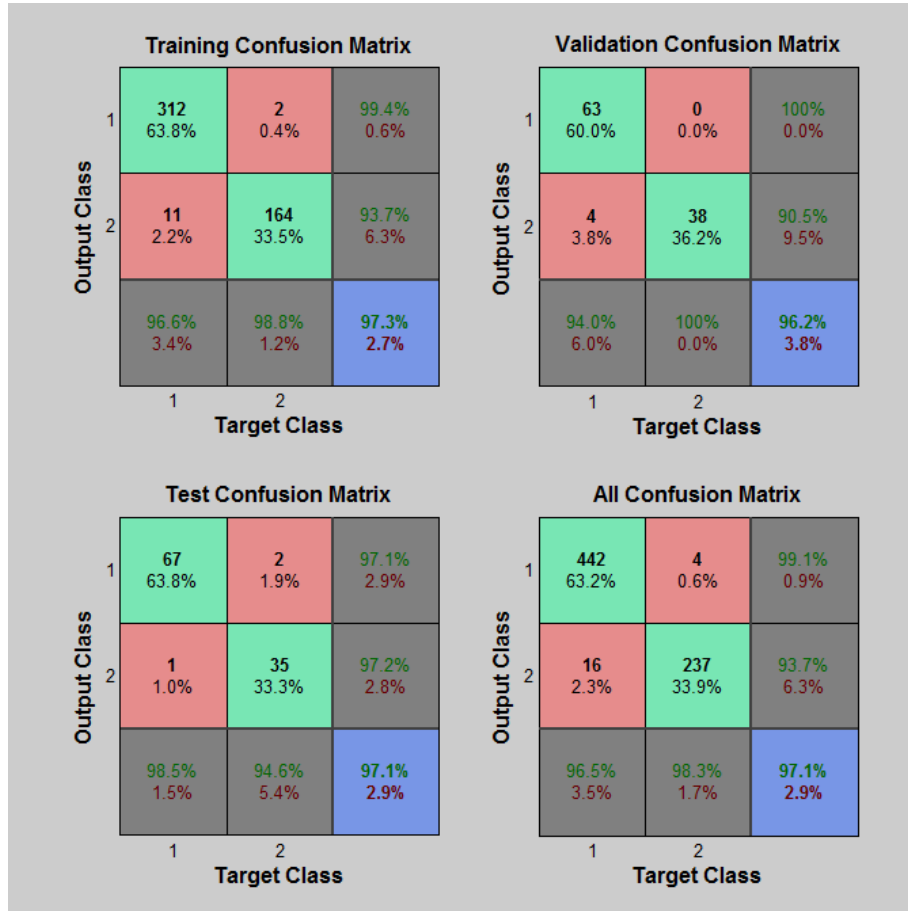
Ocena skuteczności sieci neuronowej

- W przypadku regresji:
 - RSE
 - R^2
 - F-statystyka
- W przypadku klasyfikacji:
 - Macierze pomyłek
 - Krzywe ROC

Uczenie sieci w środowisku MATLAB



Ocena sieci w środowisku MATLAB



Najważniejsze dzisiaj:

- Regresja logistyczna modeluje prawdopodobieństwo przynależności obiektu do klasy
- Do oceny klasyfikatorów mogą służyć:
 - Macierze pomyłek
 - Krzywe ROC
- Sieci neuronowe to rodzina modeli nieparametrycznych i nieinterpretowalnych
- Algorytm spadku gradientu – algorytm poszukiwania minimum funkcji
- Wsteczna propagacja błędów pozwala wykorzystać algorytm spadku gradientu do ustalenia wag w sieci neuronowej

Plan wykładów - podsumowanie

1. Wykład I – Podstawowe pojęcia, wprowadzenie do metod uczenia maszynowego
2. Wykład II – Regresja liniowa, regresja logistyczna (WIII), ocena skuteczności modelu
3. Wykład III – Sieci neuronowe, algorytm spadku gradientu, algorytm wstecznej propagacji błędów