Jakość uczenia i generalizacja

Dokładność uczenia

- Jest koncepcją miary w jakim stopniu nasza sieć nauczyła się rozwiązywać określone zadanie

- Dokładność mówi na ile nauczyliśmy się rozwiązywać zadania które już znamy, a generalizacja na ile potrafimy rozwiązywać nowe zadania podobne do tych które znamy ale mogą też być inne

Miary dokładności uczenia dla problemów klasyfikacyjnych

- Macierz konfuzji

		Rzeczywiste	
		Prawda	Fałsz
Oszacowane	Pozytywne	TP	FP
	Negatywne	FN	TN

- parametr (ang. true positive) TP
- parametr (ang. false positive) FP
- parametr (ang. true negative) TN
- parametr (ang. false negative) FN
- parametr (ang. positive) $Pos = TP + FP$
- parametr (ang. negative) $Neg = TN + FN$
- parametr (ang. true) $True = TP + TN$
- parametr (ang. false) $False = FP + FN$

Miary dokładności uczenia dla problemów klasyfikacyjnych cd.

- Dokładność (ang. accuracy)

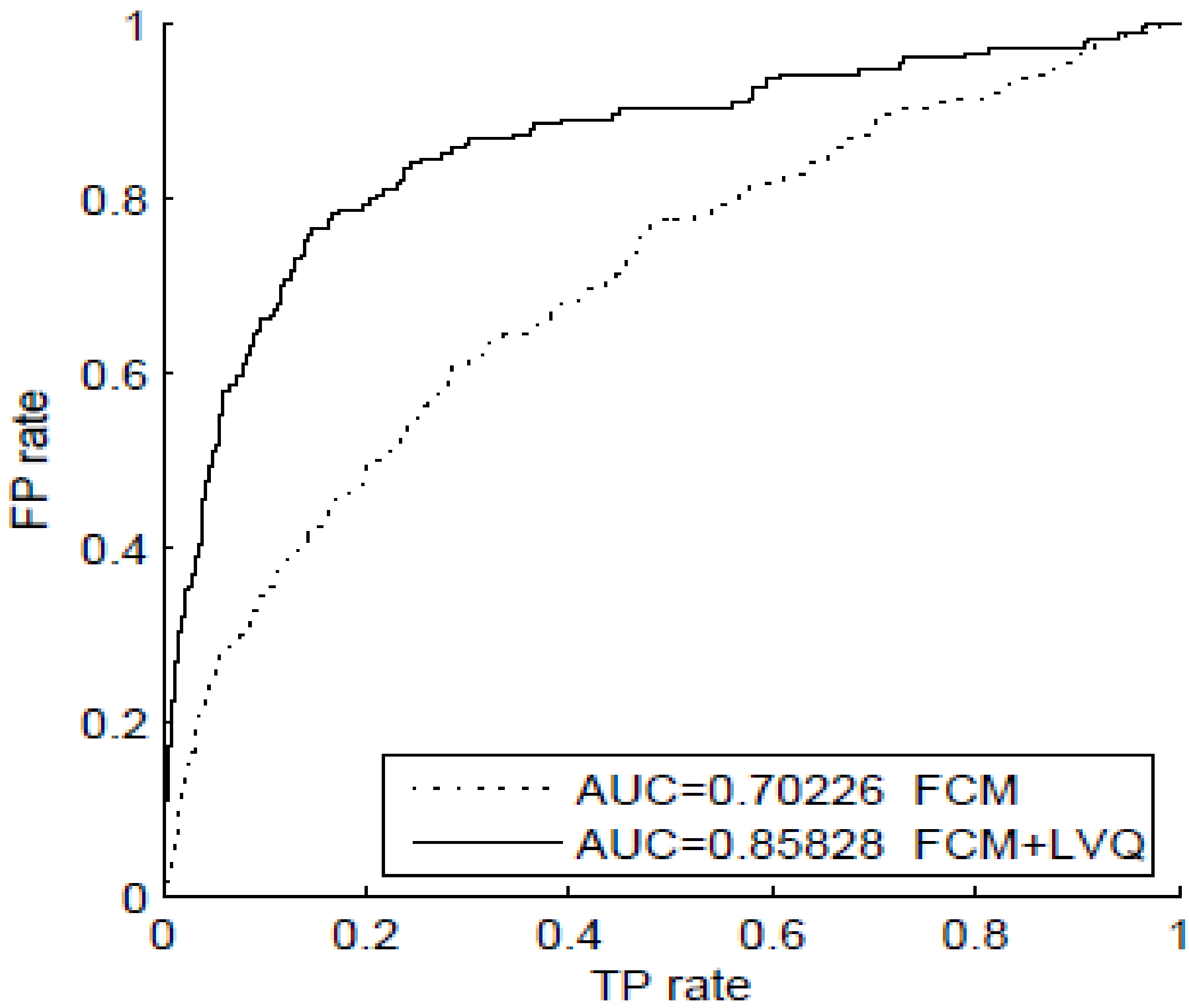
$$Err = \frac{m^{err}}{m}$$
$$Acc = 1 - Err$$

- Dokładność zbalansowana (ang. Balanced accuracy)

$$BErr = \frac{1}{c}\sum_{i=1}^{c}\left(\frac{m_i^{err}}{m_i}\right)$$
$$BAcc = 1 - BErr$$

- Czułość lub wrażliwość (ang. sensitivity, recall) $Se = \frac{TP}{TP+FN}$
- Znamienność (ang. specificity) $Sp = \frac{TN}{TN+FP}$
- Precyzja (ang. precision) $P = \frac{TP}{TP+FP}$
- Miara F_β (ang. F_β measure) $F_\beta = \frac{(\beta^2+1)TP}{(\beta^2+1)TP+FP+\beta^2 FN}$, która dla $\beta = 1$ przyjmuje postać $F_{1.0} = \frac{2TP}{2TP+FP+FN} = \frac{2Se \cdot P}{Se+P}$
- Dokładność $Acc = \frac{TP+TN}{Pos+Neg}$
- Dokładność zbalansowana $BAcc = \frac{1}{c}\left(\frac{TP}{Pos} + \frac{TN}{Neg}\right)$

Miary dokładności uczenia krzywa ROC i AUC



Miary dokładności uczenia dla problemów regresyjnych

- Błąd średniokwadratowy (MSE)

$$e = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - d_i\right)^2$$

- (RMSE)

$$e = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - d_i\right)^2}$$

- Błąd średni (ME) $e = \frac{\sum_i \left|y_i - d_i\right|}{n}$

- Znormalizowany błąd średniokwadratowy (NMSE)

$$e = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - d_i\right)^2}}{\sigma_y \sigma_d}$$

Co to jest generalizacja

- Generalizacja – zdolność do uogólniania zdobytej wiedzy

  - Jeśli wszyscy uczyliby się do egzaminu równy okres czasu to i tak będą rozbieżności w ocenach.
  - Dlaczego?

  - Gdyż jedni uczą się szybciej a inni wolniej, jedni uczą się z przyjemnością fizyki, a inni wolą uczyć się na temat malarstwa.
  - To samo zjawisko występuje w SSN i inteligencji obliczeniowej

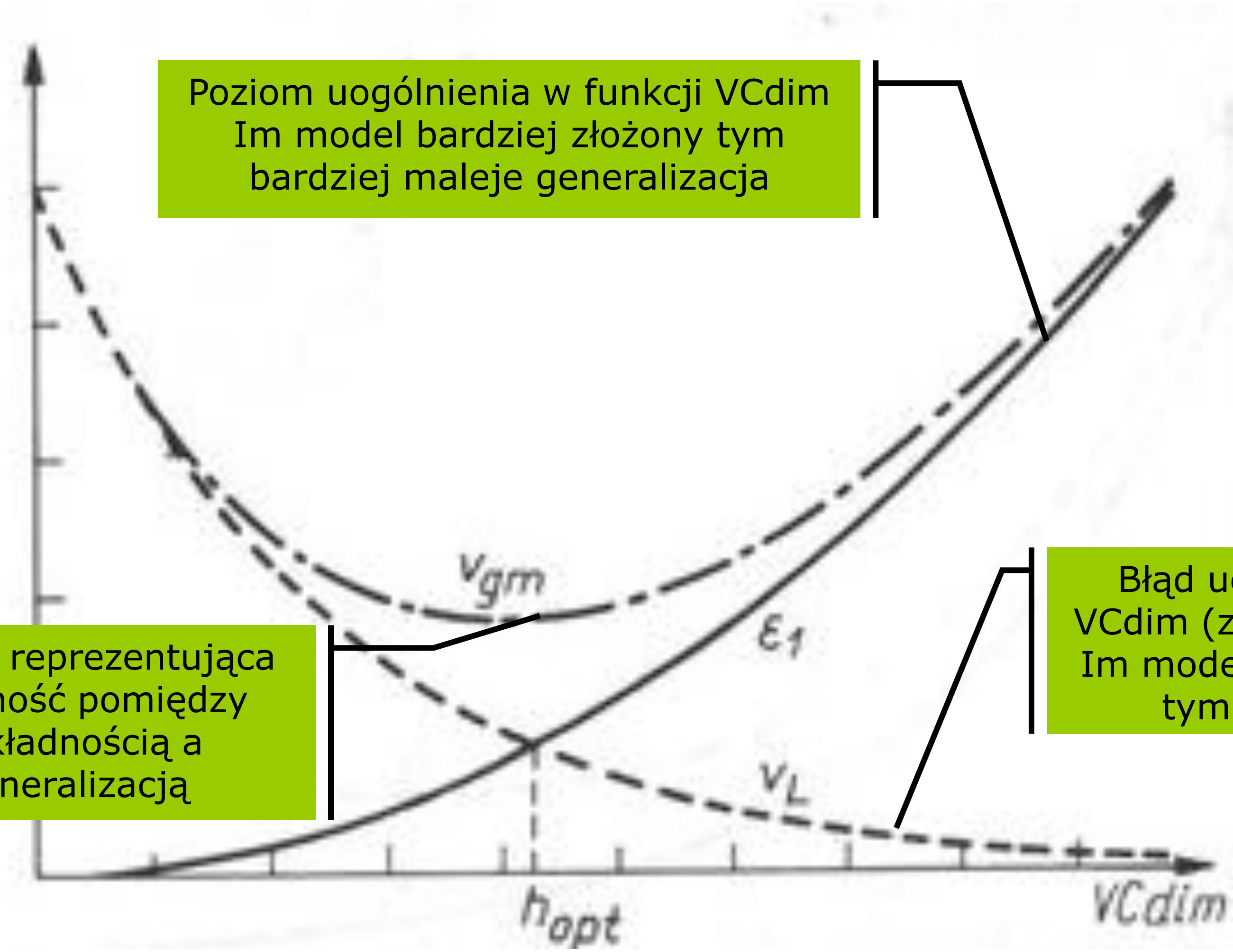
  - Generalizacja jest więc koncepcją określającą zdolnością do zastosowania zdobytej wiedzy w nowych sytuacjach, z którymi system nie spotkał się dotychczas.

Co to jest generalizacja

Generalizacja $\neq$ Dokładność

Generalizacja a dokładność

Poziom uogólnienia w funkcji VCdim
Im model bardziej złożony tym
bardziej maleje generalizacja



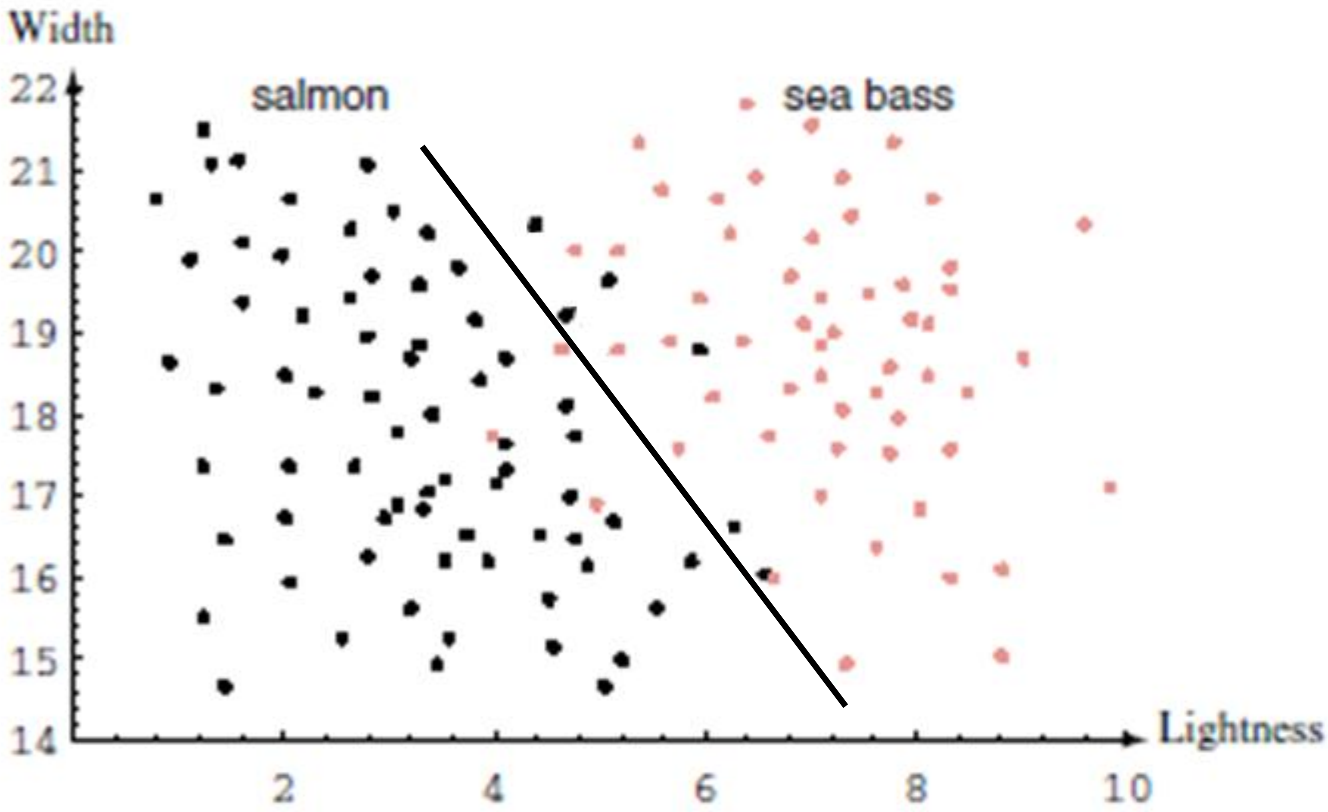
Krzywa reprezentująca
zależność pomiędzy
dokładnością a
generalizacją

Błąd uczenia w funkcji
VCdim (złożoności modelu)
Im model bardziej złożony
tym większy błąd

(*) – „Sieci neuronowe w ujęciu algorytmicznym”, Ossowski

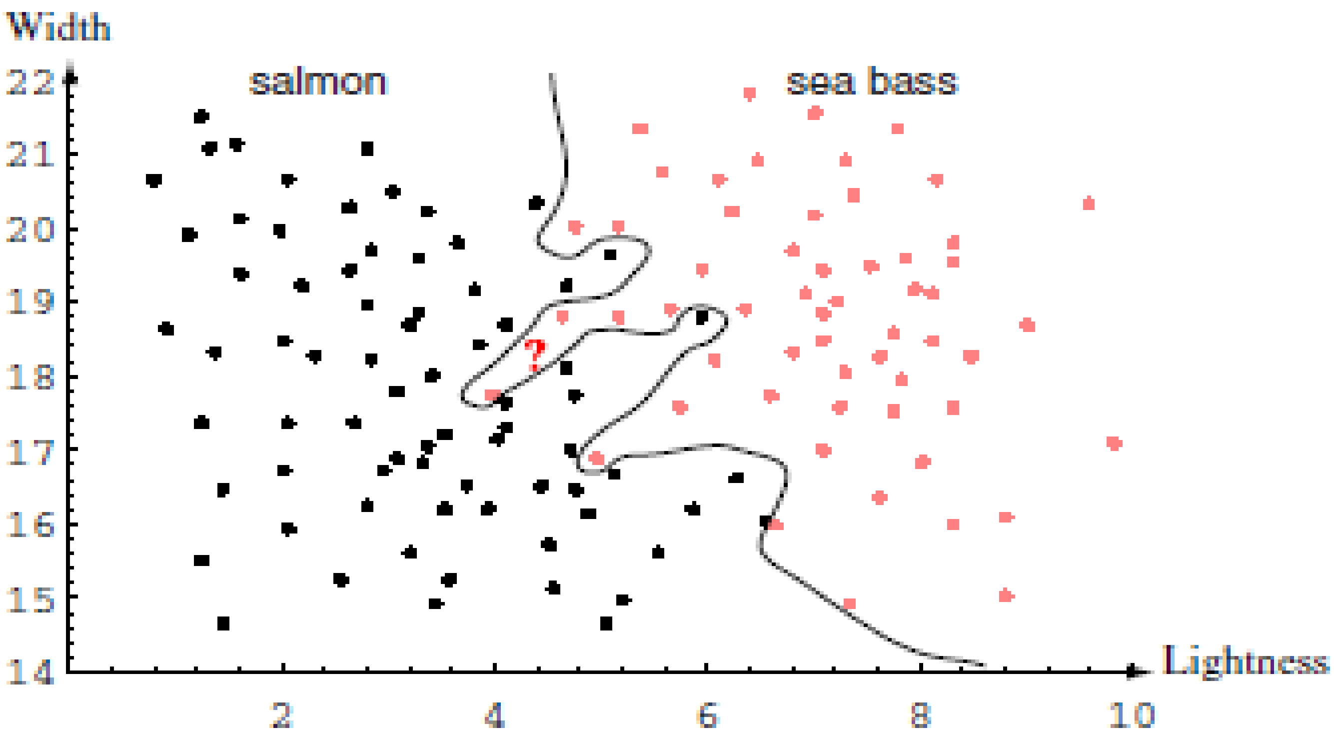
Dokładność a generalizacja

- Zbyt duża generalizacja



Dokładność a generalizacja

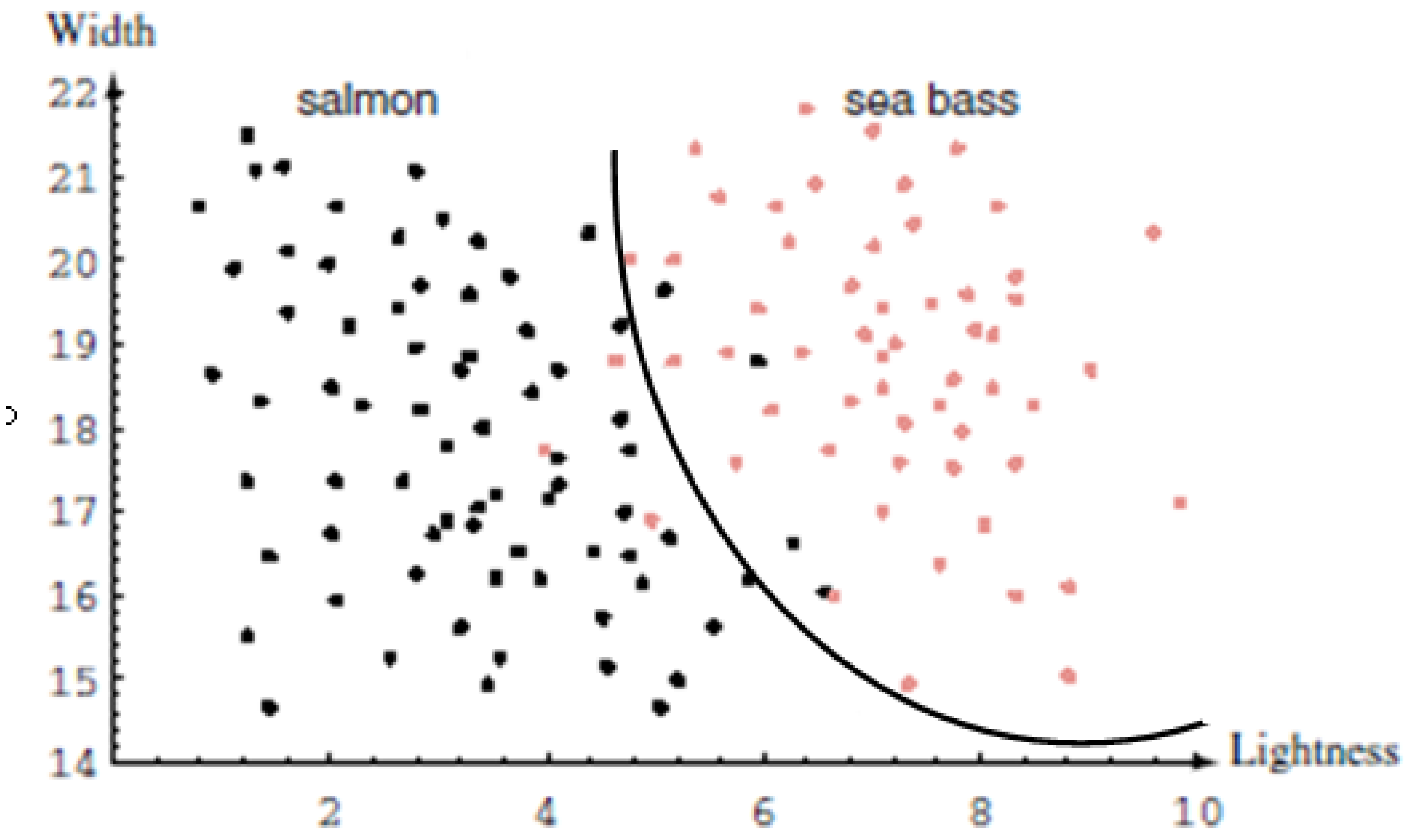
- Przeuczenie



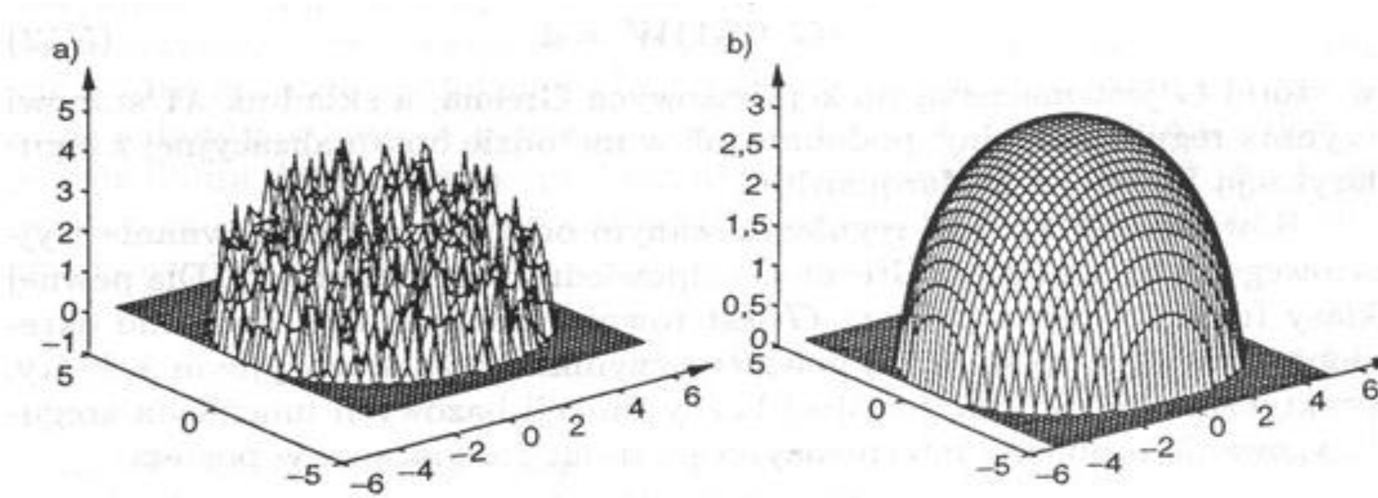
(*) – „Pattern classification” Duda, Hart, Stork

Dokładność a generalizacja

- Dobra generalizacja



Dokładność a generalizacja



(*) – „Sieci neuronowe w ujęciu algorytmicznym”, Ossowski

Problem rozmiaru sieci

- Im mamy bardziej pojemną sieć tym dłużej musimy się uczyć i tym więcej przypadków musimy przeanalizować

- Problem
  - Jeżeli mamy bardzo dużą sieć to może dojść do niedouczenia sieci, jak i do przeuczenia się danego problemu
  - Przykład
  - Jeżeli mamy do rozwiązania zadanie z fizyki aby określić drogę jaką przebył kamień spadając przez 10s to można zadanie rozwiązać:

    a) $$s = \frac{g \cdot t^2}{2}$$

    b) wyliczyć z definicji przyspieszenia iż jest to druga pochodna z drogi po czasie, robiąc odpowiednie przekształcenia i uwzględniając tarcie powietrza itp..

Metody oceny generalizacji

- Zasada minimalnej długości opisu (ang. Minimal description length) - bazuje na analizie rozmiaru zbioru oraz stopnia złożoności modelu *h* wyrażonego w bitach

$$h_{MDL} = \arg\min_{h \in H}(L_C(h) + L_C(\mathbf{X}|h))$$

Gdzie:

$L_C(h)$ - jest długością opisu modelu *h* pod warunkiem kodowania *C*

$L_C(\mathbf{X}|h)$ - jest długością opisu zbioru X pod warunkiem modelu *h* oraz kodowania *C*.

Metody oceny generalizacji cd.

- Zasada minimalizacji ryzyka strukturalnego (ang. structural risk minimization, SRM) - Idea współczynnika SRM polega na uwzględnieniu obok dokładności modelu *h* drugiego czynnika opisującego jego złożoność, tym samym wariancję modelu

$$R(h) \leq R_{emp}(h) + \sqrt{\left(\frac{VC(h)(log(2m/VC(h))+1) - log(\eta/4)}{h}\right)}$$

$R(h)$ - Rzeczywisty koszt popełnienia błędu

$R_{emp}(h)$ - Empiryczny koszt popełnienia błędu

$\mathrm{VC(h)}$ - Współczynnik Vapnika-Czervonenkisa

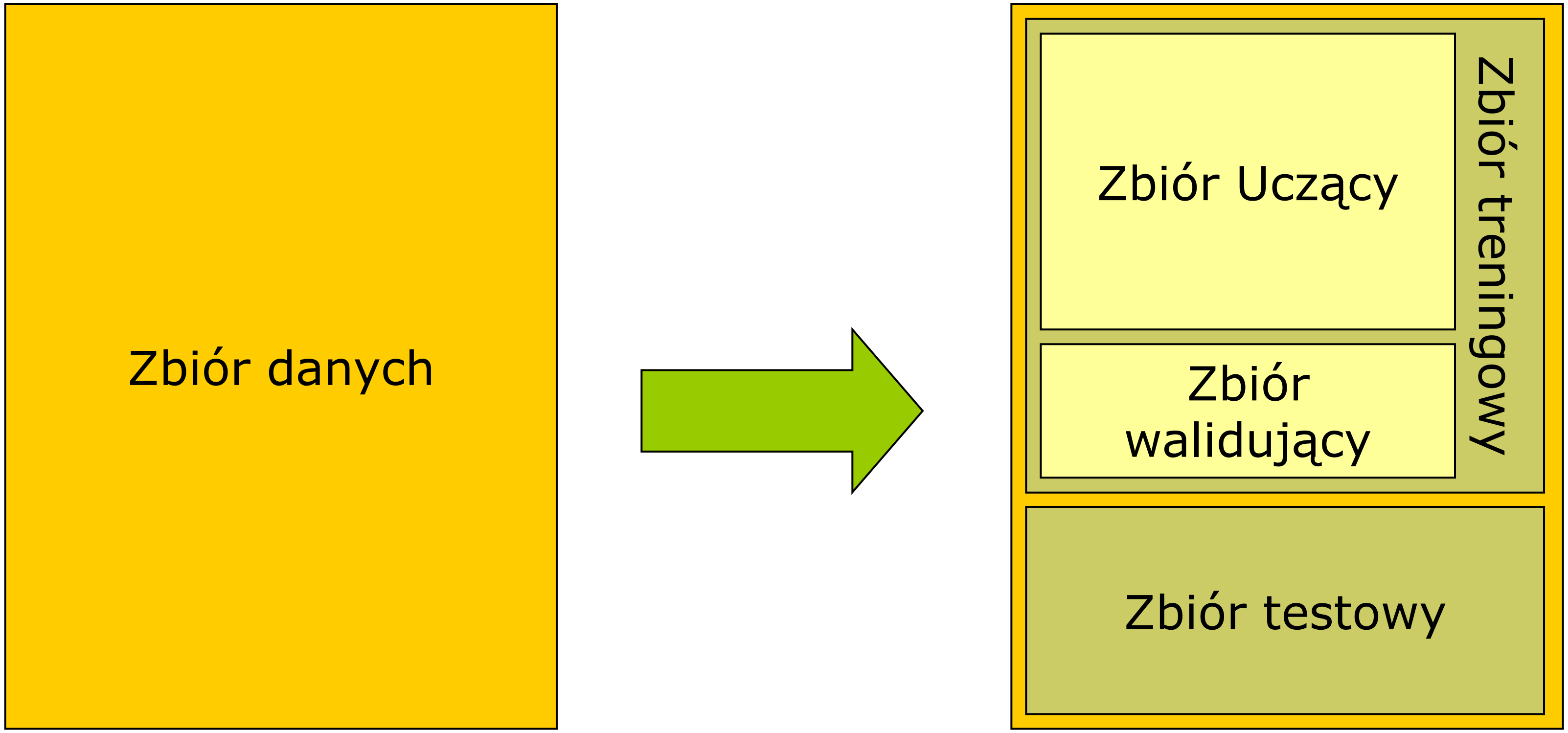
η - Pewna stała

Empiryczne metody oceny generalizacji

- Podział trening/test
- Kroswalidacja (ang cross validation) lub test krzyżowy
- Test typu jeden pozostaw
- Test typu bootstrap

Podział trening test

- Wydzielenie ze zbioru danych części na której będzie dokonywane uczenie i części na której będzie dokonywana ocena (testowanie) modelu
  - Wada – łatwo może dojść do przeuczenia modelu lub dostosowania modelu do zbioru testowego!!!

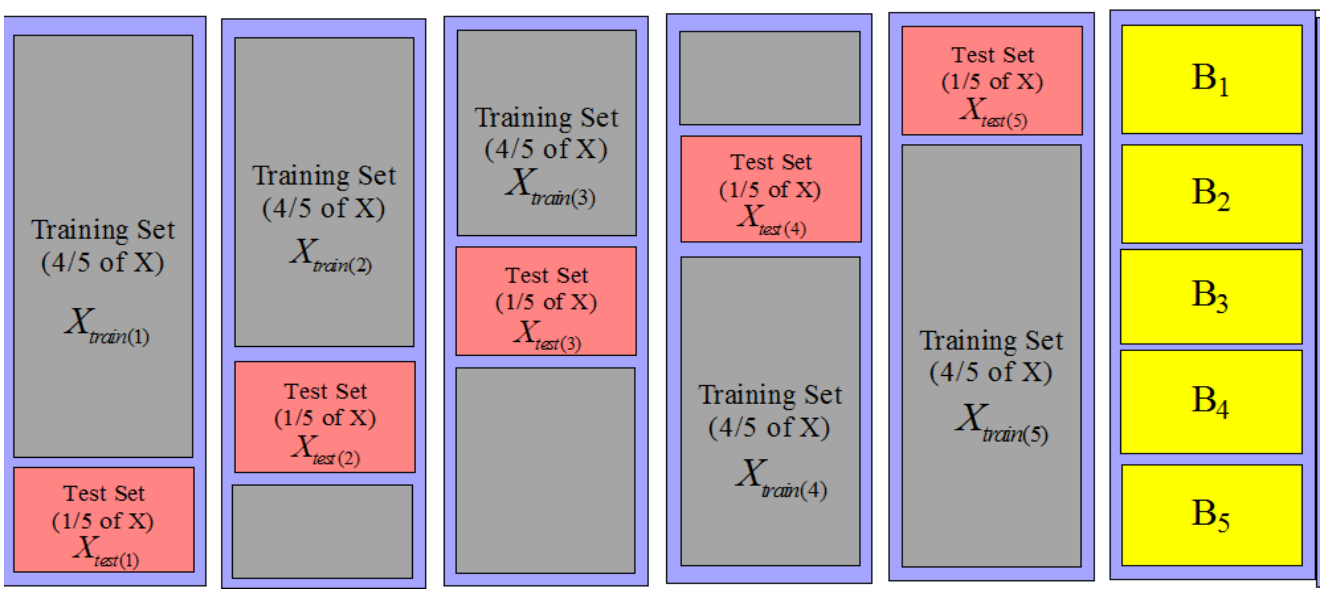


Test krzyżowy

- Podziel zbiór danych *X* na *N* równych niezależnych części R(i), i=1..N (losowanie bez zwracania)
- For i=1:N
  - Ucz model na zbiorze X / R(i)
  - Testuj model na R(i)

Dwa typy testu krzyżowego:

- normalny – losowo podział zbioru danych
- Stratyfikowany – losowy podział ale zapewniający stały rozkład punktów w poszczególnych klasach



Jeden pozostaw

- Test jeden pozostaw (ang. Leave one out) jest szczególną wersją testu krzyżowego, gdzie zawsze tylko 1 wektor pozostawiany jest do testowania, a proces uczenia odbywa się M razy, gdzie M – liczba wektorów w zbiorze uczącym

- Zalety – test nie obciążony
- Wady – duża złożoność obliczeniowa (znane są różne uproszczenia testu LOO)

Bootstrap (by Efron)

Mając zbiór X o M wektorach

- Ze zbioru danych X wylosuj ze zwracaniem M wektorów aby powstał zbiór R_j
- Naucz model H_j na zbiorze R_j i wyznacz błąd modelu na tym zbiorze RE_j
- Wyznacz błąd XE_j modelu H_j na całym zbiorze danych X
- Wyznacz optymizm $opt_j = XE_j - Re_j$
- Powtórz kroki 1:4 J krotnie
- Wyznacz średni optymizm $\underline{opt}$
- Wyznacz błąd na zbiorze uczącym E (trening modelu i jego testowanie odbywa się na zbiorze X)
- Na tej podstawie estymatę błędu generalizacji (e) *wynosi*
  $e = E + \underline{opt}$