Weryfikacja i walidacja

Piotr Guzik, Lucjan Janowski, Krzysztof Rusek

November 13, 2017

Outline

Jak idzie?

Outline

Narzędzia do przykładów

Najważniejsze cechy wielomianów.

Narzędzia do przykładów

Najważniejsze cechy wielomianów.

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} = 1 - \frac{\text{Zmienność niewyjaśniona}}{\text{Całkowita zmienność}}$$

Narzędzia do przykładów

Najważniejsze cechy wielomianów.

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} = 1 - \frac{\text{Zmienność niewyjaśniona}}{\text{Całkowita zmienność}}$$

$$\text{MSE} = \sum(y_i - \hat{y}_i)^2 = \text{Zmienność niewyjaśniona}$$

The Bias-Variance Trade-Off

Zakładamy, że rzeczywistość opisuje równanie:

$$y = f(x) + \epsilon$$

The Bias-Variance Trade-Off

Zakładamy, że rzeczywistość opisuje równanie:

$$y = f(x) + \epsilon$$

Wtedy model można opisać równaniem:

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

The Bias-Variance Trade-Off

Zakładamy, że rzeczywistość opisuje równanie:

$$y = f(x) + \epsilon$$

Wtedy model można opisać równaniem:

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

Licząc błąd posłużymy się równaniem:

$$E((y - \hat{y})^2)$$

Błąd

$$E((y-\hat{y})^2) = E((f(x)-\hat{f}(x))^2) + E(\epsilon^2) + E(2(f(x)-\hat{f}(x))\epsilon)$$

Błąd

$$E((y-\hat{y})^2) = E((f(x)-\hat{f}(x))^2) + E(\epsilon^2) + E(2(f(x)-\hat{f}(x))\epsilon)$$

Ponieważ błąd jest niezależny otrzymujemy:

$$E((y-\hat{y})^2) = E((f(x)-\hat{f}(x))^2) + E(\epsilon^2) + 0$$

Gdzie $E(\epsilon^2)$ jest nieredukowalnym błędem wynikającym z ograniczeń systemu pomiaru.

Błąd

$$E((y-\hat{y})^2) = E((f(x)-\hat{f}(x))^2) + E(\epsilon^2) + E(2(f(x)-\hat{f}(x))\epsilon)$$

Ponieważ błąd jest niezależny otrzymujemy:

$$E((y-\hat{y})^2) = E((f(x)-\hat{f}(x))^2) + E(\epsilon^2) + 0$$

Gdzie $E(\epsilon^2)$ jest nieredukowalnym błędem wynikającym z ograniczeń systemu pomiaru. Błąd wynikający z różnicy pomiędzy rzeczywistością $f(x)$ oraz modelem $\hat{f}(x)$ można rozpisać:

$$E((f(x)-\hat{f}(x))^2) = Var(\hat{f}(x)) + [\mathrm{Bias}(\hat{f}(x))]^2$$

$Var(\hat{f}(x))$

Jest to podatność funkcji na zmianę danych. Czyli jak bardzo zmienią się wartości dla zmienionych danych.

Kod, sekcja 1.

$[\mathrm{Bias}(\hat{f}(x))]^2$

Jest to możliwość odwzorowywania zmienności danych. Jak dokładnie zmienną rzeczywistość może odwzorować dana funkcja.

Kod, sekcja 2

$Var(\hat{f}(x))$

Generalnie algorytmy o większej liczbie parametrów mają większą wariancję, jednak liczba parametrów to nie wszystko.

Wariancja jest zawsze większa dla bardziej złożonych algorytmów, ale dla dużej ilości danych może okazać się, że wyniki złożonych algorytmów są porównywalne z prostymi. Wpływ wariancji maleje wraz z ilością danych.

$[\mathrm{Bias}(\hat{f}(x))]^2$

Proste modele często oddają najważniejsze właściwości rzeczywistości.

$[\mathrm{Bias}(\hat{f}(x))]^2$

Proste modele często oddają najważniejsze właściwości rzeczywistości.
David Kahneman: Stabilne małżeństwo

$$\text{liczba stosunków} - \text{liczba kłótni} > 0$$

$[\mathrm{Bias}(\hat{f}(x))]^2$

Proste modele często oddają najważniejsze właściwości rzeczywistości.
David Kahneman: Stabilne małżeństwo

$$\text{liczba stosunków} - \text{liczba kłótni} > 0$$

Proste modele można rozumieć, obronić w sądzie, czy przedyskutować z praktykiem.

$[\mathrm{Bias}(\hat{f}(x))]^2$

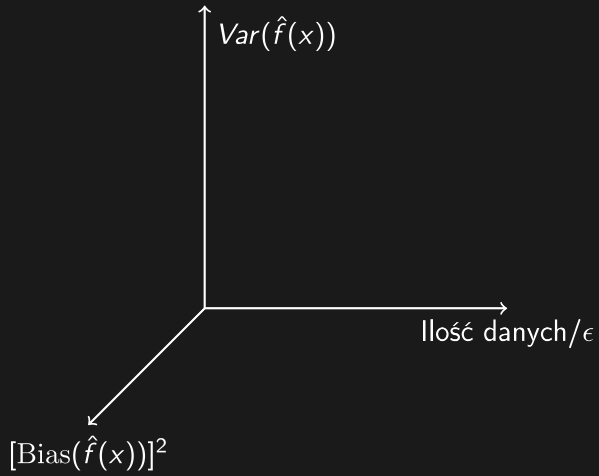
Proste modele często oddają najważniejsze właściwości rzeczywistości.
David Kahneman: Stabilne małżeństwo

$$\text{liczba stosunków} - \text{liczba kłótni} > 0$$

Proste modele można rozumieć, obronić w sądzie, czy przedyskutować z praktykiem.

Wydaje się, że rzeczywistość jest skomplikowana.

Trzy wymiary błędu



Kod sekcja 3

Różne modele

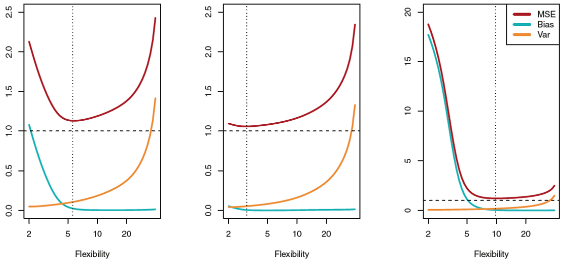


FIGURE 2.12. *Squared bias (blue curve), variance (orange curve), Var(ϵ) (dashed line), and test MSE (red curve) for the three data sets in Figures 2.9–2.11. The vertical dotted line indicates the flexibility level corresponding to the smallest test MSE.*

Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani
An Introduction to Statistical Learning, with Applications in R
Springer, 2015, strona 36

Kod sekcja 4

Outline

Walidacja i weryfikacja

walidacja (za słownikiem PWN)

weryfikacja (za słownikiem PWN)

Walidacja i weryfikacja

walidacja (za słownikiem PWN)

- ogół czynności mających na celu zbadanie odpowiedniości, trafności lub dokładności czegoś

weryfikacja (za słownikiem PWN)

Walidacja i weryfikacja

walidacja (za słownikiem PWN)

- ogół czynności mających na celu zbadanie odpowiedniości, trafności lub dokładności czegoś

weryfikacja (za słownikiem PWN)

- sprawdzenie prawdziwości, przydatności lub prawidłowości czegoś
- ocena pracownika i sprawdzenie jego przydatności na danym stanowisku

Walidacja i weryfikacja

walidacja

- Koniczne dla każdego proponowanego rozwiązania
- Sprawdzenie czy algorytm faktycznie działa
- Podstawa każdego procesu tworzenia algorytmu
- Najczęściej stosowana jest walidacja krosowa
- Pozwala na wybieranie lepszych algorytmów ew. dobieranie odpowiednich parametrów algorytmów

Walidacja krosowa

Set 1	Set 2	Set 3	Set 4	Set 5
Set 6	Set 7	Set 8	Set 9	Set 10

Zbiór treningowy

Zbiór testowy = MSE1

Walidacja krosowa

Set 1	Set 2	Set 3	Set 4	Set 5
Set 6	Set 7	Set 8	Set 9	Set 10

Zbiór treningowy

Zbiór testowy = MSE2

Walidacja krosowa

Set 1	Set 2	Set 3	Set 4	Set 5
Set 6	Set 7	Set 8	Set 9	Set 10

Zbiór treningowy

Zbiór testowy = MSE10

Walidacja krosowa wynik

$$\mathrm{MSE} = \frac{1}{10}\sum_{i=1}^{10} \mathrm{MSE}i$$

Walidacja krosowa wynik

$$\mathrm{MSE} = \frac{1}{10}\sum_{i=1}^{10}\mathrm{MSE}i$$

$$R^2 = \frac{1}{10}\sum_{i=1}^{10}R_i^2$$

Walidacja i weryfikacja

weryfikacja

- Nie ma wpływu na sam algorytm
- Profesjonalna weryfikacja tworzona jest przez niezależną organizację
- Pozwala na ocenę algorytmu dla danego zastosowania
- Może mieć problem z obszarem działania algorytmu

Walidacja krosowa

Zbiór na którym stworzono algorytm

Zbiór weryfikacyjny

Zakres działania algorytmu

Algorytm rozpoznawania jedzenia. Co mamy na talerzu.

Zakres działania algorytmu

Algorytm rozpoznawania jedzenia. Co mamy na talerzu. Jeżeli nauczmy go rozpoznawać jedzenie z Indii, to co będzie z:

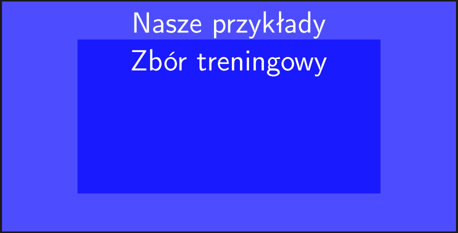
Zakres działania algorytmu

Algorytm rozpoznawania jedzenia. Co mamy na talerzu.

Zbór treningowy

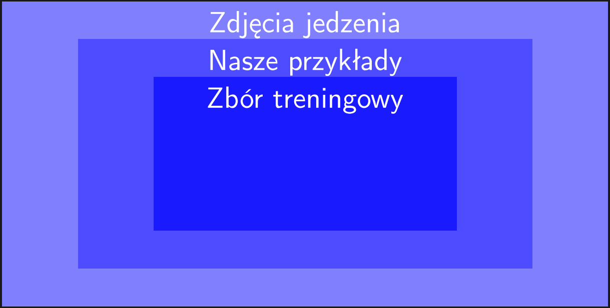
Zakres działania algorytmu

Algorytm rozpoznawania jedzenia. Co mamy na talerzu.



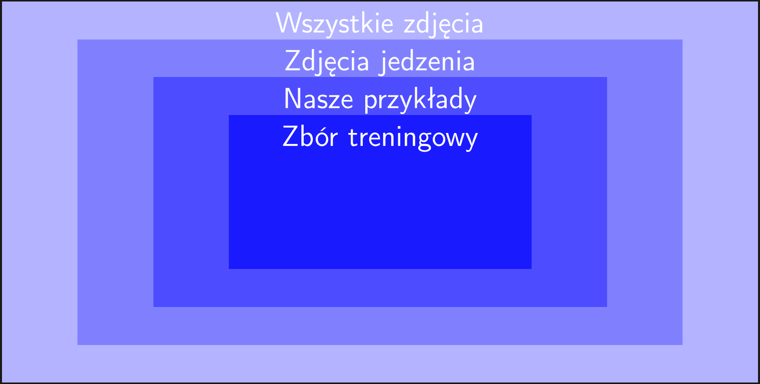
Zakres działania algorytmu

Algorytm rozpoznawania jedzenia. Co mamy na talerzu.



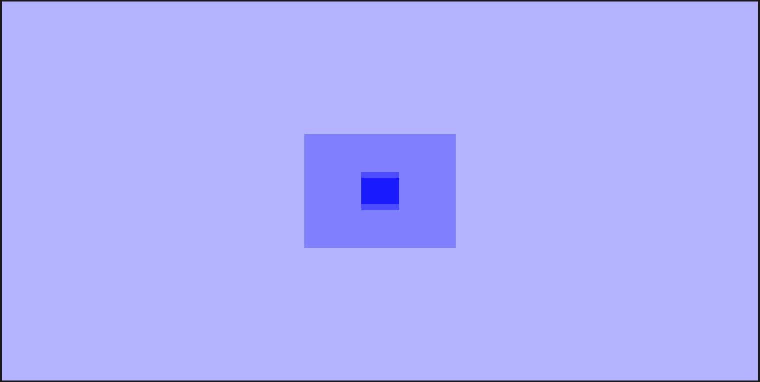
Zakres działania algorytmu

Algorytm rozpoznawania jedzenia. Co mamy na talerzu.



Zakres działania algorytmu, w skali

Algorytm rozpoznawania jedzenia. Co mamy na talerzu.



Walidacja krosowa

Dzielimy zbiór testowy na n zbiorów i na $n-1$ ćwiczymy model a na n-tym walidujemy wynik. Najczęściej $n = 10$. Czynność powtarzamy wiele razy.

Trzeba pamiętać o zakresie działania algorytmu i co czyni przykłady istotnie różnymi! Podział ma być po kategoriach, niekoniecznie sztukach.

Kod sekcja 5

Outline

Odcinanie brzegów

Chcemy obciąć 5% brzegowych wartości dla każdej ze zmiennych. Zmiennych jest n i wszystkie są od 0 do 1. Jaką część zbioru odcinamy?

Odcinanie brzegów

Chcemy obciąć 5% brzegowych wartości dla każdej ze zmiennych. Zmiennych jest n i wszystkie są od 0 do 1. Jaką część zbioru odcinamy?

$$n = 1$$

Odcinanie brzegów

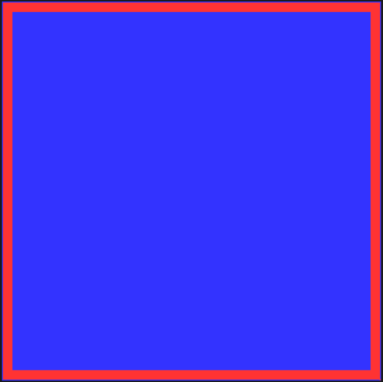
Chcemy obciąć 5% brzegowych wartości dla każdej ze zmiennych. Zmiennych jest n i wszystkie są od 0 do 1. Jaką część zbioru odcinamy?

$$n = 1$$

Dla 1000 pomiarów po odcięciu brzegu, przy założeniu równomierności pomiarów, otrzymujemy $1000 * 0.95 = 950$ pomiarów.

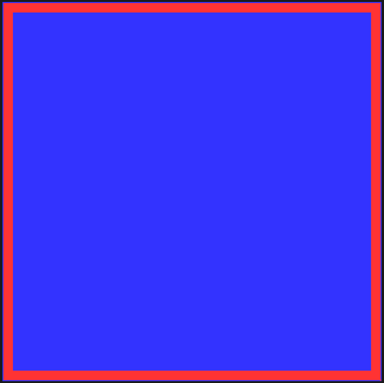
Odcinanie brzegów

$n = 2$



Odcinanie brzegów

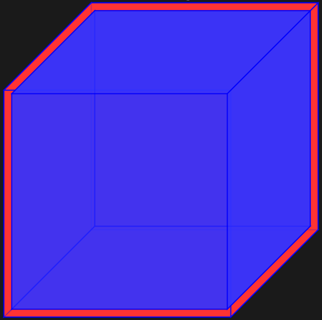
$n = 2$



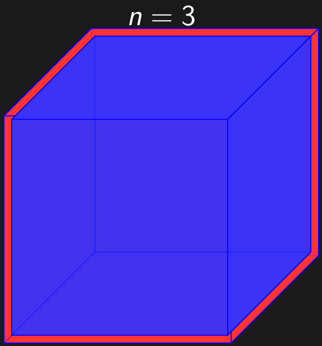
Dla 1000 pomiarów po odcięciu brzegu otrzymujemy $1000 * 0.95^2 \sim 903$ pomiarów.

Odcinanie brzegów

$n = 3$



Odcinanie brzegów



Dla 1000 pomiarów po odcięciu brzegu otrzymujemy $1000 * 0.95^3 \sim 857$ pomiarów.

Odcinanie brzegów

$n = 5$
Dla 1000 pomiarów po odcięciu brzegu otrzymujemy
$1000 * 0.95^5 \sim 774$ pomiarów.

Odcinanie brzegów

$n = 5$
Dla 1000 pomiarów po odcięciu brzegu otrzymujemy
$1000 * 0.95^5 \sim 774$ pomiarów.
$n = 50$
Dla 1000 pomiarów po odcięciu brzegu otrzymujemy
$1000 * 0.95^{50} \sim 77$ pomiarów.

Odcinanie brzegów

$n = 5$

Dla 1000 pomiarów po odcięciu brzegu otrzymujemy $1000 * 0.95^5 \sim 774$ pomiarów.

$n = 50$

Dla 1000 pomiarów po odcięciu brzegu otrzymujemy $1000 * 0.95^{50} \sim 77$ pomiarów.

$n = ?$

Kiedy po odcięciu "brzegów" otrzymamy 0 pomiarów czyli $1000 * 0.95^n < 0.5$?

Odcinanie brzegów

$n = 5$

Dla 1000 pomiarów po odcięciu brzegu otrzymujemy

$1000 * 0.95^5 \sim 774$ pomiarów.

$n = 50$

Dla 1000 pomiarów po odcięciu brzegu otrzymujemy

$1000 * 0.95^{50} \sim 77$ pomiarów.

$n = ?$

Kiedy po odcięciu "brzegów" otrzymamy 0 pomiarów czyli

$1000 * 0.95^n < 0.5$?

$n \geq 149$

Inne spojrzenie

Jeżeli spodziewamy się liniowej zależności, 5- 7 punktów pomiarowych powinno wystarczyć. Dla dwóch niezależnych zmiennych daje to 25 - 49 pomiarów.

Inne spojrzenie

Jeżeli spodziewamy się liniowej zależności, 5- 7 punktów pomiarowych powinno wystarczyć. Dla dwóch niezależnych zmiennych daje to 25 - 49 pomiarów.

Dla czterech zmiennych to 625 - 2401.

Inne spojrzenie

Jeżeli spodziewamy się liniowej zależności, 5- 7 punktów pomiarowych powinno wystarczyć. Dla dwóch niezależnych zmiennych daje to 25 - 49 pomiarów.

Dla czterech zmiennych to 625 - 2401.

Dla dziesięciu zmiennych to 9 765 625 - 282 475 249.