

Uczenie oparte o pamiętanie

ML by ML

Mateusz Lango

Zakład Inteligentnych Systemów Wspomagania Decyzji
Wydział Informatyki i Telekomunikacji
Politechnika Poznańska

19 listopada 2020

Uczenie oparte o pamiętanie

- Intuicja 1: Pamiętanie to ważna część uczenia się
 - Intuicja 2: Przykłady podobne w cechach powinny być podobne w zmiennej decyzyjnej
- ⇒ Najbliżsi sąsiedzi x (wg. jakiejś funkcji odległości) powinni być tego samego typu
- $$y = f(x) \approx f(N(x))$$
- ⇒ Predykcja to agregacja wartości w sąsiedztwie $N(x)$

Problem

Czy inne poznane techniki też podejmują decyzję zgodnie z powyższą zasadą?

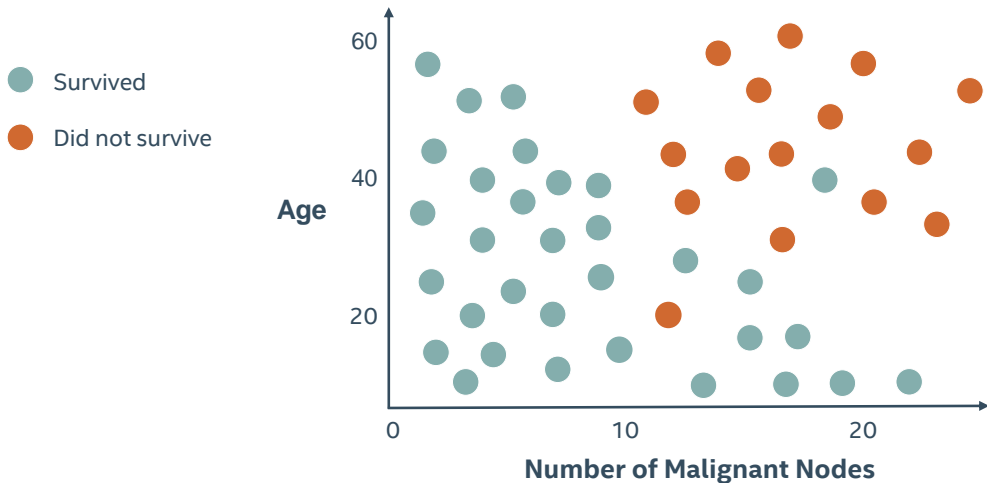
Uczenie oparte o pamiętanie

- Intuicja 1: Pamiętanie to ważna część uczenia się
 - Intuicja 2: Przykłady podobne w cechach powinny być podobne w zmiennej decyzyjnej
- ⇒ Najbliżsi sąsiedzi x (wg. jakiejś funkcji odległości) powinni być tego samego typu
- $$y = f(x) \approx f(N(x))$$
- ⇒ Predykcja to agregacja wartości w sąsiedztwie $N(x)$

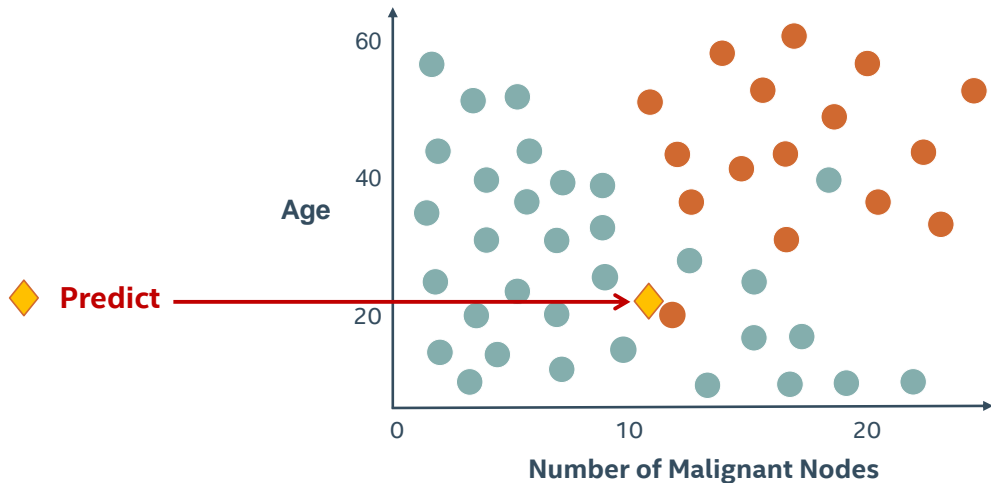
Problem

Czy inne poznane techniki też podejmują decyzję zgodnie z powyższą zasadą?

K NEAREST NEIGHBORS CLASSIFICATION



K NEAREST NEIGHBORS CLASSIFICATION



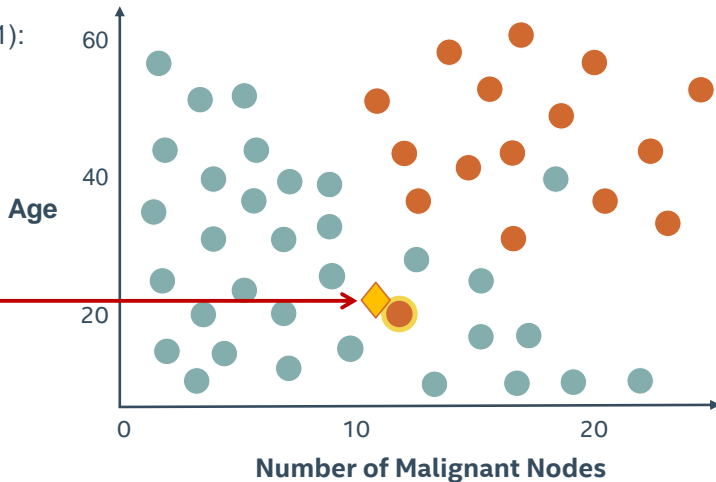
K NEAREST NEIGHBORS CLASSIFICATION

Neighbor Count ($K = 1$):

● 0

● 1

◆ Predict



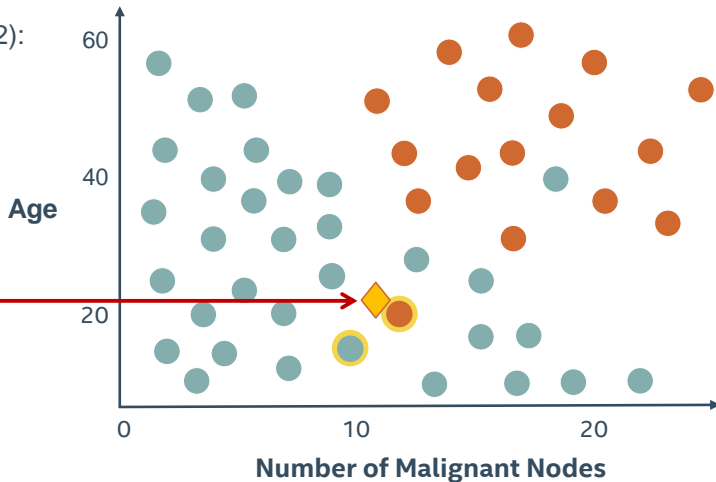
K NEAREST NEIGHBORS CLASSIFICATION

Neighbor Count ($K = 2$):

● 1

● 1

◆ Predict



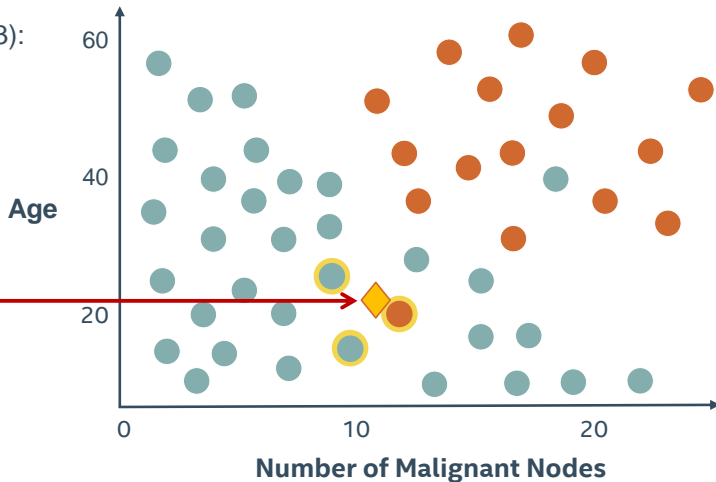
K NEAREST NEIGHBORS CLASSIFICATION

Neighbor Count ($K = 3$):

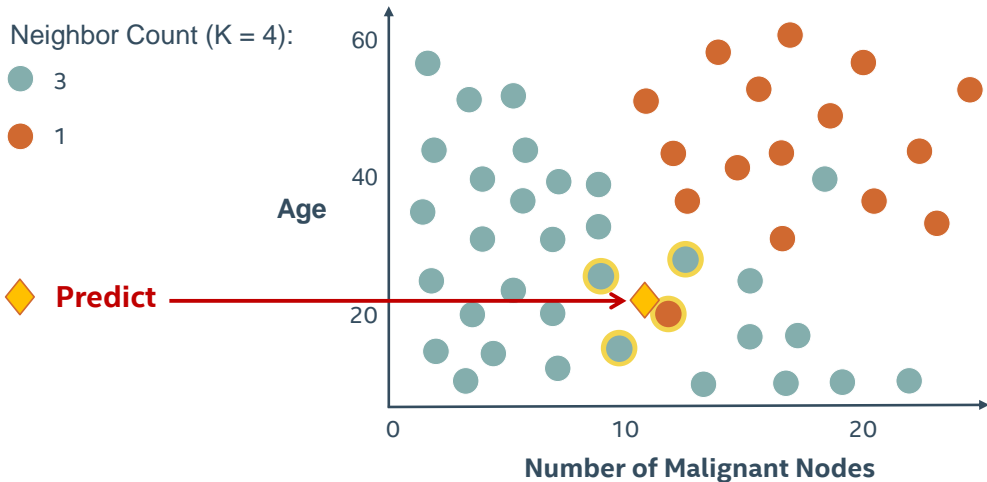
● 2

● 1

◆ Predict



K NEAREST NEIGHBORS CLASSIFICATION



Klasyfikator k najbliższych sąsiadów

- Faza uczenia
 - ① Nic. (zakładając że przykłady uczące są zapisane w pamięci – jeśli nie to zapisz)
- Faza predykcji
 - ① Oblicz odległość pomiędzy przykładem testowym a wszystkimi przykładami w zbiorze uczącym
 - ② Zidentyfikuj k najbliższych elementów (sortowanie)
 - ③ Użyj etykiet znalezionych przykładów do określenia etykiety przykładu testowego (głosowanie większościowe)
- Co potrzebuję ustawić?
 - ① Trzeba określić liczbę najbliższych sąsiadów k
 - ② Trzeba zdefiniować funkcję odległości pomiędzy przykładami

Klasyfikator k najbliższych sąsiadów

- Faza uczenia
 - ① Nic. (zakładając że przykłady uczące są zapisane w pamięci – jeśli nie to zapisz)
- Faza predykcji
 - ① Oblicz odległość pomiędzy przykładem testowym a wszystkimi przykładami w zbiorze uczącym
 - ② Zidentyfikuj k najbliższych elementów (sortowanie)
 - ③ Użyj etykiet znalezionych przykładów do określenia etykiety przykładu testowego (głosowanie większościowe)
- Co potrzebuję ustawić?
 - ① Trzeba określić liczbę najbliższych sąsiadów k
 - ② Trzeba zdefiniować funkcję odległości pomiędzy przykładami

Klasyfikator k najbliższych sąsiadów

- Faza uczenia
 - ① Nic. (zakładając że przykłady uczące są zapisane w pamięci – jeśli nie to zapisz)
- Faza predykcji
 - ① Oblicz odległość pomiędzy przykładem testowym a wszystkimi przykładami w zbiorze uczącym
 - ② Zidentyfikuj k najbliższych elementów (sortowanie)
 - ③ Użyj etykiet znalezionych przykładów do określenia etykiety przykładu testowego (głosowanie większościowe)
- Co potrzebuję ustawić?
 - ① Trzeba określić liczbę najbliższych sąsiadów k
 - ② Trzeba zdefiniować funkcję odległości pomiędzy przykładami

Klasyfikator k najbliższych sąsiadów – pytania

Problem

Jaka jest złożoność obliczeniowa uczenia i predykcji?

Problem

Co się dzieje jak mamy więcej niż k najbliższych sąsiadów (ta sama odległość)?

Problem

Czy relacja „najbliższego sąsiada” jest symetryczna?

Problem

Czy mogę uzyskać z kNN wyjście probabilistyczne?

Problem

Czy mogę uzyskać z kNN regresor?

Klasyfikator k najbliższych sąsiadów – pytania

Problem

Jaka jest złożoność obliczeniowa uczenia i predykcji?

Problem

Co się dzieje jak mamy więcej niż k najbliższych sąsiadów (ta sama odległość)?

Problem

Czy relacja „najbliższego sąsiada” jest symetryczna?

Problem

Czy mogę uzyskać z kNN wyjście probabilistyczne?

Problem

Czy mogę uzyskać z kNN regresor?

Klasyfikator k najbliższych sąsiadów – pytania

Problem

Jaka jest złożoność obliczeniowa uczenia i predykcji?

Problem

Co się dzieje jak mamy więcej niż k najbliższych sąsiadów (ta sama odległość)?

Problem

Czy relacja „najbliższego sąsiada” jest symetryczna?

Problem

Czy mogę uzyskać z kNN wyjście probabilistyczne?

Problem

Czy mogę uzyskać z kNN regresor?

Klasyfikator k najbliższych sąsiadów – pytania

Problem

Jaka jest złożoność obliczeniowa uczenia i predykcji?

Problem

Co się dzieje jak mamy więcej niż k najbliższych sąsiadów (ta sama odległość)?

Problem

Czy relacja „najbliższego sąsiada” jest symetryczna?

Problem

Czy mogę uzyskać z kNN wyjście probabilistyczne?

Problem

Czy mogę uzyskać z kNN regresor?

Klasyfikator k najbliższych sąsiadów – pytania

Problem

Jaka jest złożoność obliczeniowa uczenia i predykcji?

Problem

Co się dzieje jak mamy więcej niż k najbliższych sąsiadów (ta sama odległość)?

Problem

Czy relacja „najbliższego sąsiada” jest symetryczna?

Problem

Czy mogę uzyskać z kNN wyjście probabilistyczne?

Problem

Czy mogę uzyskać z kNN regresor?

Klasyfikator k najbliższych sąsiadów – wybór k

Problem

Jak wybrać k ? Czy mamy jakieś intuicje? Jak działa klasyfikator kNN dla $k = N$?

Klasyfikator k najbliższych sąsiadów – wybór k

Problem

Jak wybrać k ? Czy mamy jakieś intuicje? Jak działa klasyfikator kNN dla $k = N$?

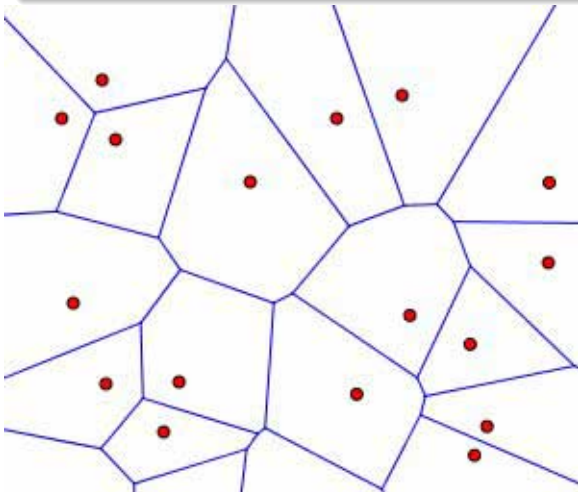


Diagram pokazujący regiony, których najbliższym sąsiadem jest jeden z punktów to diagram Woronoja.

Klasyfikator k najbliższych sąsiadów – wybór miary odległości

Klasyfikator uzyskuje uogólnianie na nowe przykłady poprzez stosowanie funkcji odległości – to w niej zaszyte są umiejętności klasyfikatora! \Rightarrow absolutnie kluczowy parametr modelu!

- Własne funkcje odległości wykorzystujące wiedzę dziedzinową
- Kilka klasycznych:
 - Atrybuty nominalne?
 - Atrybuty ciągłe?

Klasyfikator k najbliższych sąsiadów – wybór miary odległości

Klasyfikator uzyskuje uogólnianie na nowe przykłady poprzez stosowanie funkcji odległości – to w niej zaszyte są umiejętności klasyfikatora! \Rightarrow absolutnie kluczowy parametr modelu!

- Własne funkcje odległości wykorzystujące wiedzę dziedzinową
- Kilka klasycznych:
 - Atrybuty nominalne?
 - Atrybuty ciągłe?

Klasyfikator k najbliższych sąsiadów – wybór miary odległości

Klasyfikator uzyskuje uogólnianie na nowe przykłady poprzez stosowanie funkcji odległości – to w niej zaszyte są umiejętności klasyfikatora! \Rightarrow absolutnie kluczowy parametr modelu!

- Własne funkcje odległości wykorzystujące wiedzę dziedzinową
- Kilka klasycznych:
 - Atrybuty nominalne?
 - Atrybuty ciągłe?

Value Difference Metric (VDM) – sprytne miara dla nominalnych

$$vdm_a(x, y) = \sum_{c \in C} |P(c|x_a) - P(c|y_a)|$$

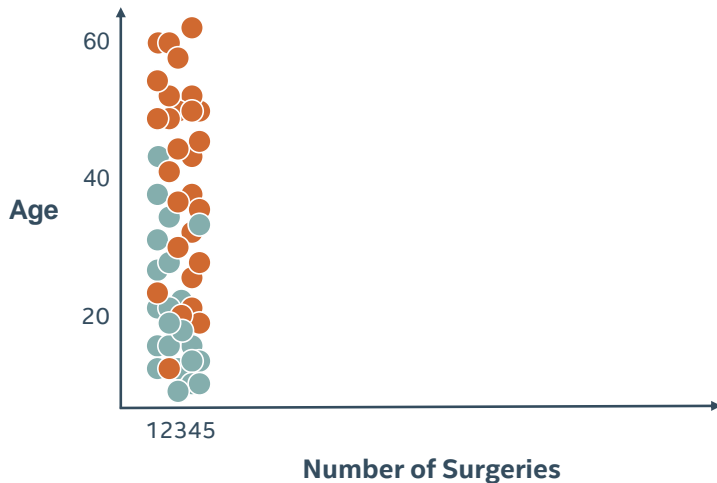
Problem

Cecha 1	Cecha 2	Klasa
P	C	+
N	C	+
Z	C	-
N	R	+
Z	R	-

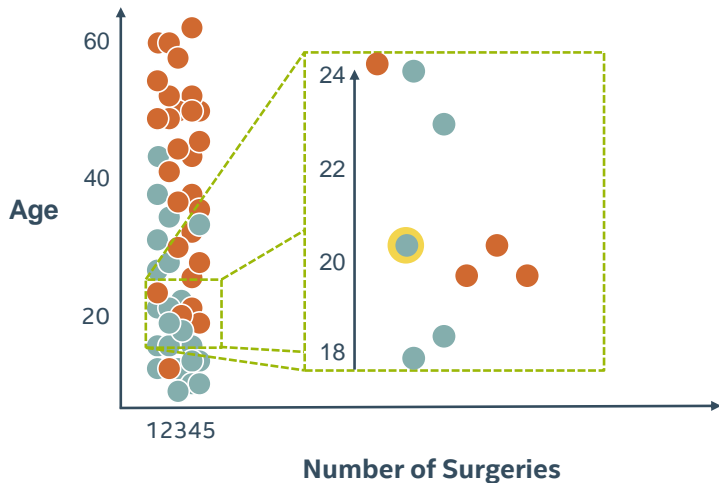
Oblicz VDM pomiędzy przykładem o cechach (N,C) a przykładem (P,R)

Miary dla mieszanych danych: HVDM, Gower distance

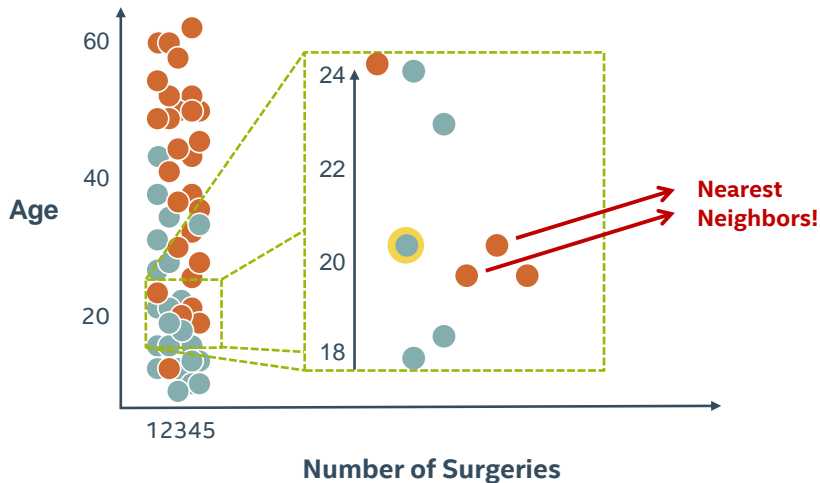
SCALE IS IMPORTANT FOR DISTANCE MEASUREMENT



SCALE IS IMPORTANT FOR DISTANCE MEASUREMENT

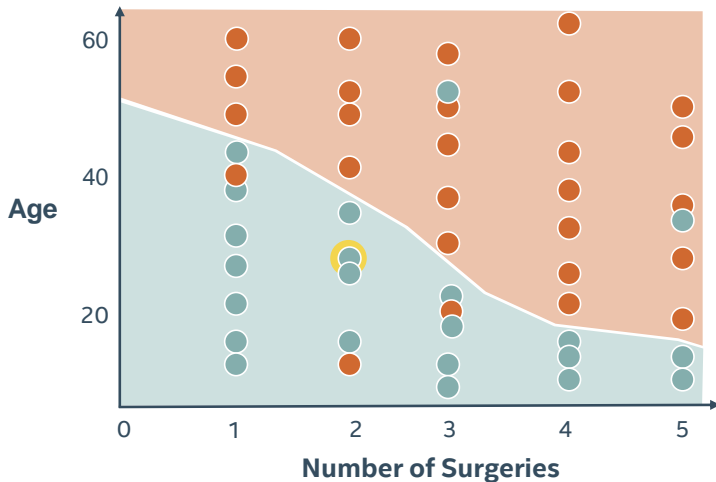


SCALE IS IMPORTANT FOR DISTANCE MEASUREMENT



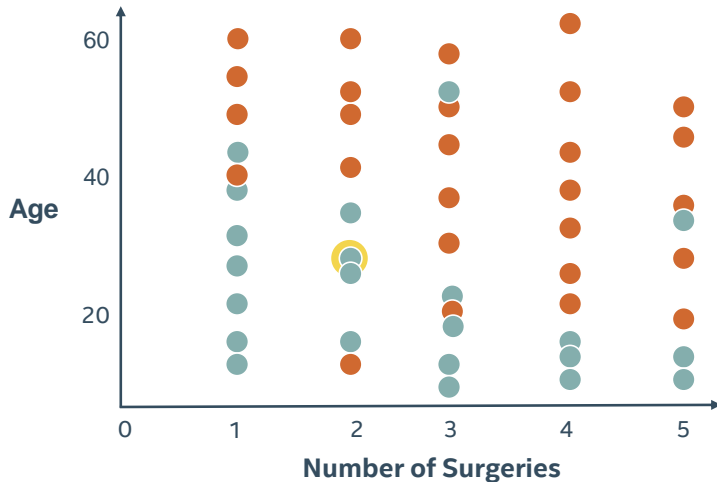
SCALE IS IMPORTANT FOR DISTANCE MEASUREMENT

"Feature Scaling"



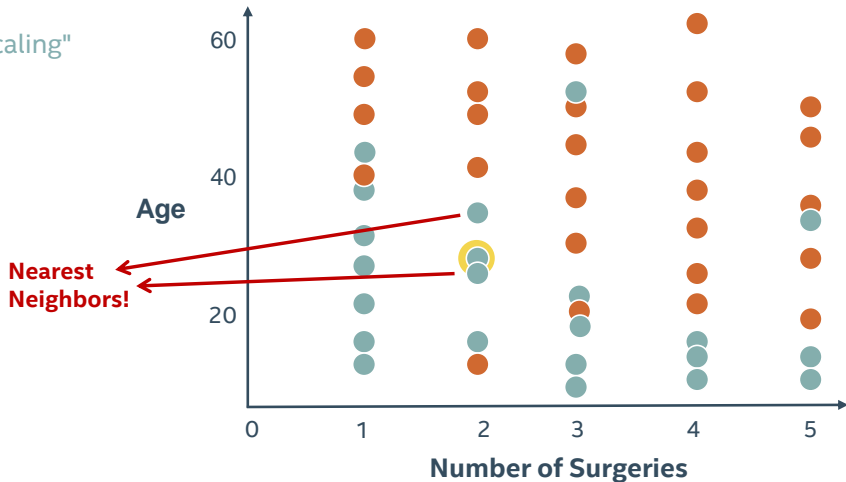
SCALE IS IMPORTANT FOR DISTANCE MEASUREMENT

"Feature Scaling"



SCALE IS IMPORTANT FOR DISTANCE MEASUREMENT

"Feature Scaling"



Własność uniwersalności

Dotychczas poznane klasyfikatory mają swoje ograniczenia tj. nie potrafią modelować wszystkich możliwych funkcji. Jak to wygląda z kNN?

Problem

Podaj przykłady par takich funkcji i metod.

Theorem

Jeżeli dane (X, Y) pochodzą z próby losowej prostej, a

$$Y = f(X) + \text{losowy szum}$$

*gdzie f jest **dowolnie złożoną** funkcją, a parametr k jest wybrany tak aby $\frac{k}{n} \rightarrow 0$ i $k \rightarrow \infty$ to*

$$\lim_{n \rightarrow \infty} \mathbb{E}[|\hat{f}(X) - f(X)|] = 0$$

Własność uniwersalności

Dotychczas poznane klasyfikatory mają swoje ograniczenia tj. nie potrafią modelować wszystkich możliwych funkcji. Jak to wygląda z kNN?

Problem

Podaj przykłady par takich funkcji i metod.

Theorem

Jeżeli dane (X, Y) pochodzą z próby losowej prostej, a

$$Y = f(X) + \text{losowy szum}$$

*gdzie f jest **dowolnie złożoną** funkcją, a parametr k jest wybrany tak aby $\frac{k}{n} \rightarrow 0$ i $k \rightarrow \infty$ to*

$$\lim_{n \rightarrow \infty} \mathbb{E}[|\hat{f}(X) - f(X)|] = 0$$

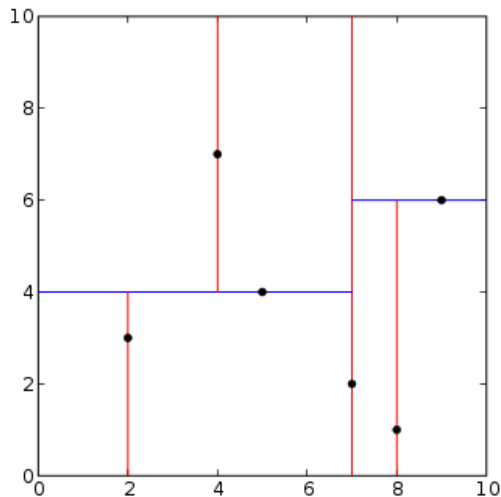
Klasyfikator kNN – wady i zalety

Zanim przejdziemy dalej:

Problem

Jakie są mocne i słabe strony klasyfikatora kNN?

Klasyfikator kNN – problem ze złożonością obliczeniową



Działanie klasyfikator kNN można przyspieszyć przy pomocy odpowiednich struktury jak np. kd-tree.

Średni czas poszukiwania: $O(\log n)$

Maksymalny czas poszukiwania: $O(n)$

(Podejścia statystyczne, przybliżone – znacznie szybsze)

Klasyfikator kNN – problem ze złożonością pamięciową

Przy okazji: uczenie przyrostowe

Uczenie przyrostowe zajmuje się tworzeniem algorytmów, które nie mają dostępu do całego zbioru danych i mają dostępną na raz tylko jedną (losową) instancję uczącą.

Condensed nearest neighbor – dodaj etap uczenia się, który polega na wyborze podzbioru instancji potrzebnych do dalszej klasyfikacji.

Przechodzimy do analizy działania kNN...

Kończymy omawianie klasyfikatora kNN poprzez przedstawienie uproszczonej analizy teoretycznej jednego z aspektów jego działania.

Czy są do tego momentu jakieś pytania?

- Optymistycznie założyliśmy, że k najbliższych sąsiadów jest blisko klasyfikowanego przykładu
- Wynika to z intuicji „jeśli przykłady są podobne to mają podobne klasy”
- Czy tak jednak jest naprawdę? Jaki jest rozmiar kuli obejmującej k najbliższych sąsiadów?

Przechodzimy do analizy działania kNN...

Kończymy omawianie klasyfikatora kNN poprzez przedstawienie uproszczonej analizy teoretycznej jednego z aspektów jego działania.

Czy są do tego momentu jakieś pytania?

- Optymistycznie założyliśmy, że k najbliższych sąsiadów jest blisko klasyfikowanego przykładu
- Wynika to z intuicji „jeśli przykłady są podobne to mają podobne klasy”
- Czy tak jednak jest naprawdę? Jaki jest rozmiar kuli obejmującej k najbliższych sąsiadów?

Czy rzeczywiście NN jest blisko x ?

- Załóżmy, że kula $B(x)$ jest najmniejszą kulą zawierającą kNN przykładu x o promieniu $r(x)$.
- Ile wynosi (zgrubnie) promień tej kuli?

$$r(x) \approx \sqrt[d]{\frac{k}{n} \cdot \frac{1}{p_X(x)}}$$

Problem

Jakie płyną wnioski z tego wzoru?

Czy rzeczywiście NN jest blisko x ?

- Załóżmy, że kula $B(x)$ jest najmniejszą kulą zawierającą kNN przykładu x o promieniu $r(x)$.
- Ile wynosi (zgrubnie) promień tej kuli?

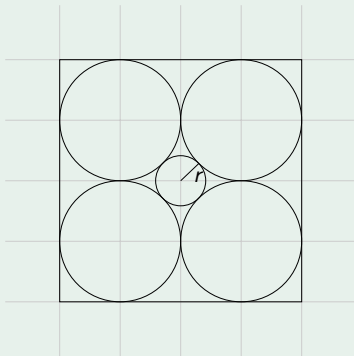
$$r(x) \approx \sqrt[d]{\frac{k}{n} \cdot \frac{1}{p_X(x)}}$$

Problem

Jakie płyną wnioski z tego wzoru?

Kłątwa wymiarowości

Problem



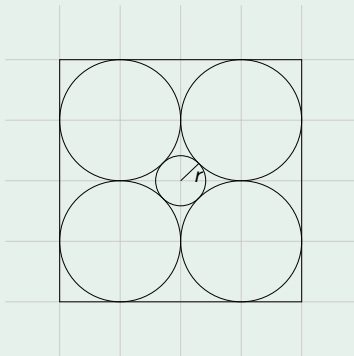
Ile wynosi r ?

Problem

Ile wynosi r dla analogicznej kuli w 3 wymiarach?

Kłątwa wymiarowości

Problem



Ile wynosi r ?

Problem

Ile wynosi r dla analogicznej kuli w 3 wymiarach?

Klątwa wymiarowości

- Dla $d = 2$ promień koła w środku to $\sqrt{2} - 1 \approx 0.41$
- Dla $d = 3$ promień kuli w środku to $\sqrt{3} - 1 \approx 0.73$
- I jest to ogólna zależność tj. dla przypadku d wymiarowego promień wynosi $r = \sqrt{d} - 1$
- Dla $d = 4$ promień hiperkuli w środku to $\sqrt{4} - 1 = 1$!!!
- Dla $d = 16$ promień hiperkuli w środku to $\sqrt{16} - 1 = 3$!!!

Problem

Implikacje dla ML?

Konkretne zjawiska i problemy kłątwy wymiarowości \Rightarrow Ćwiczenia
Iskierka nadziei: „prawdziwa” wymiarowość danych \Rightarrow Ćwiczenia

Klątwa wymiarowości

- Dla $d = 2$ promień koła w środku to $\sqrt{2} - 1 \approx 0.41$
- Dla $d = 3$ promień kuli w środku to $\sqrt{3} - 1 \approx 0.73$
- I jest to ogólna zależność tj. dla przypadku d wymiarowego promień wynosi $r = \sqrt{d} - 1$
- Dla $d = 4$ promień hiperkuli w środku to $\sqrt{4} - 1 = 1$!!!
- Dla $d = 16$ promień hiperkuli w środku to $\sqrt{16} - 1 = 3$!!!

Problem

Implikacje dla ML?

Konkretne zjawiska i problemy klątwy wymiarowości \Rightarrow Ćwiczenia
Iskierka nadziei: „prawdziwa” wymiarowość danych \Rightarrow Ćwiczenia

Klątwa wymiarowości

- Dla $d = 2$ promień koła w środku to $\sqrt{2} - 1 \approx 0.41$
- Dla $d = 3$ promień kuli w środku to $\sqrt{3} - 1 \approx 0.73$
- I jest to ogólna zależność tj. dla przypadku d wymiarowego promień wynosi $r = \sqrt{d} - 1$
- Dla $d = 4$ promień hiperkuli w środku to $\sqrt{4} - 1 = 1$!!!
- Dla $d = 16$ promień hiperkuli w środku to $\sqrt{16} - 1 = 3$!!!

Problem

Implikacje dla ML?

Konkretne zjawiska i problemy klątwy wymiarowości \Rightarrow Ćwiczenia
Iskierka nadziei: „prawdziwa” wymiarowość danych \Rightarrow Ćwiczenia

Dekompozycja obciążenie-wariancja

Theorem

Jeżeli dane (X, Y) pochodzą z próby losowej prostej, a

$$Y = f(X) + \text{losowy szum}$$

(i przy dodatkowych założeniach, w szczególności zakładamy^a, że w małym otoczeniu $|f(x) - f(y)| \approx |x - y|$)

$$\mathbb{E}[(\hat{f}(X) - f(X))^2] \lesssim \frac{\sigma_Y^2}{k} + \left(\frac{k}{n}\right)^{2/d}$$

^aNa nasze potrzeby, pełne dowody i twierdzenia znajdziecie w literaturze

Widzimy się za tydzień!