Zeszyty Naukowe Wydziału Elektrotechniki i Automatyki Politechniki Gdańskiej Nr 71

XXX Seminarium

ZASTOSOWANIE KOMPUTERÓW W NAUCE I TECHNICE 2020

Oddział Gdański PTETiS

doi: 10.32016/1.71.02

ELEMENTY UCZENIA MASZYNOWEGO NA ZAJĘCIACH MATEMATYKI

Agnieszka BARTŁOMIEJCZYK[1], Dawid PTACH[2], Marcin WATA[3]

1. Politechnika Gdańska, Wydział Fizyki Technicznej i Matematyki Stosowanej
   tel.: 58 347 1347 e-mail: agnbartl@pg.edu.pl
2. Politechnika Gdańska, Wydział Fizyki Technicznej i Matematyki Stosowanej
   tel.: 665 589 906 e-mail: s181495@student.pg.edu.pl
3. Uniwersytet Mikołaja Kopernika, Wydział Nauk Ekonomicznych i Zarządzania
   tel.: 56 611 4606 e-mail: marcin.wata@umk.pl

Streszczenie: W artykule omówiono związki między matematyką kursową a wybranymi zagadnieniami związanymi z uczeniem maszynowym. Pokazano w jaki sposób proste operacje na macierzach pomagają serwisom VOD w rekomendacji tytułów filmowych zgodnych z zainteresowaniami użytkowników na podstawie ich wcześniejszych wyborów. Zaprezentowano również uproszczoną wersję algorytmu regresji wielorakiej stosowaną do wyceny nieruchomości oraz wspomniano o zastosowaniu sieci neuronowych w problemach klasyfikacyjnych.

Słowa kluczowe: uczenie maszynowe, regresja wieloraka, problemy klasyfikacyjne.

1. WSTĘP

Sytuacja epidemiczna na świecie wymusiła przejście na tryb zdalnego nauczania w szkołach i na uczelniach wyższych. Komunikacja synchroniczna z sali wykładowych i ćwiczeniowych przeniosła się do webinariów na różnych platformach. Do dotychczasowych problemów w nauczaniu matematyki, postrzeganej jako przedmiot trudny, dołączył problem weryfikacji efektów kształcenia. Dlatego tak ważne jest wzbudzenie ciekawości studentów i pobudzenie wewnętrznej motywacji do głębszego zrozumienia poruszanych na zajęciach zagadnień.

Jedną z możliwości motywowania studentów do nauki jest pokazanie wykorzystania podstawowych elementów matematyki, z którymi studenci zapoznają się już na pierwszym roku, w praktycznym działaniu. W artykule [1] pokazaliśmy proste zastosowanie pojęć matematycznych, nauczanych w toku studiów, do analizy modeli epidemiologicznych, podkreślając konieczność użycia komputerów do efektywnego wykonywania obliczeń.

W prezentowanym artykule opisujemy kolejne przykłady, z obszaru uczenia maszynowego (ang. *machine learning*), umożliwiające połączenie elementarnej matematyki z obliczeniami numerycznymi. Dodatkową motywacją dla studentów do zapoznania się z tą tematyką mogą być niedawne sukcesy naukowców z Politechniki Gdańskiej, którzy opracowali skuteczny algorytm uczenia maszynowego do oceny złośliwości guzów nerek na podstawie zdjęcia tomografii komputerowej, [2].

2. INFORMACJE OGÓLNE

2.1. Krótka historia uczenia maszynowego

Narodziny uczenia maszynowego datuje się na lata 50. XX w. Początki tej technologii są powiązane z warcabami. Amerykański informatyk Arthur Samuel uważał, że nauczenie komputera grania w gry może być znaczące w rozwiązywaniu innych bardziej złożonych problemów. Krokiem milowym w historii uczenia maszynowego był system Dendral, którego zadaniem było ustalenie struktury molekularnej nieznanych związków organicznych na podstawie analizy widm spektroskopowych. Prace nad technologią uczenia maszynowego przyspieszyły na początku lat 90. Ogromnym sukcesem w tej dziedzinie była wygrana maszyny Deep Blue w szachy z ówczesnym szachowym mistrzem świata Garri Kasparowem.

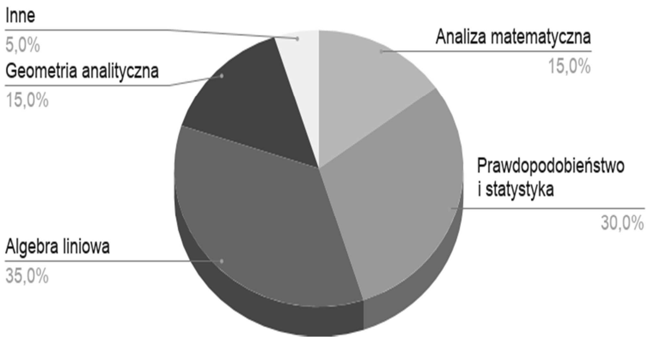
Obecnie technologia wykorzystująca uczenie maszynowe stosowana jest w wielu urządzeniach, na przykład w trakcie logowania do smartfonów, w edytorze Word do automatycznej zamiany równań pisanych odręcznie na druk, do rozpoznawania głosu i obrazu. Na podstawie naszego zachowania w Internecie personalizowane są także reklamy oraz propozycje filmów. Co więcej, większość użytkowników Internetu może nie mieć nawet świadomości, że są źródłem danych dla maszyn w procesie uczenia maszynowego.

2.2. Studenckie spojrzenie na uczenie się matematyki

Często zdarza się, że studenci pierwszych semestrów studiów zdobywają podstawy wiedzy matematycznej bez większej świadomości jej zastosowania w innych dziedzinach. Myślą, że wektory, macierze i inne obiekty czy pojęcia matematyczne żyją własnym życiem, są jakąś abstrakcyjną, alternatywną rzeczywistością. Po czasie odkrywają, że matematyka jest silnie obecna nie tylko w fizyce, ale również w chemii, biologii, a nawet w medycynie i prawie. Matematyka jest także potrzebna by zrozumieć technologię uczenia maszynowego.

Uczenie maszynowe bazuje na danych, które często reprezentowane są jako wektory i macierze. Dobre zrozumienie operacji wykonywanych na danych wymaga znajomości podstaw algebry liniowej. Przydatne są również inne działy matematyki, takie jak analiza matematyczna (zastosowanie gradientu w minimalizacji funkcji kosztu), rachunek prawdopodobieństwa i statystyka matematyczna (miara poziomu niepewności danych oraz niezawodności algorytmu).

Znaczenie poszczególnych działów matematyki w uczeniu maszynowym opracowane na podstawie [3] zostało zaprezentowane na rysunku 1.



Rys. 1. Znaczenie działów matematyki w uczeniu maszynowym

3. PRZYKŁADY UCZENIA MASZYNOWEGO

3.1. Co łączy macierze, uczenie maszynowe i serwis Netflix?

Uczenie maszynowe wykorzystywane jest przez popularne serwisy internetowe VOD (ang. *Video On Demand*), których przykładem jest Netfix. W czasie rejestracji do serwisu następuje inicjalizacja procesu uczenia. Netflix proponuje nam kilka wybranych tytułów filmowych prosząc o wskazanie tych, którymi jesteśmy zainteresowani. Następnie na podstawie takich czynników jak godziny logowania do systemu, ilości czasu spędzanego na oglądaniu wybranych programów czy filmów oraz naszych wyszukiwań, Netflix szacuje prawdopodobieństwo, że dany tytuł przyciągnie naszą uwagę, [4]. Oznacza to, że serwis nie tylko analizuje nasze wyszukiwania, ale również uczy się naszych zachowań. Co więcej, system próbuje nas zakwalifikować do grupy użytkowników o podobnych preferencjach dzięki czemu optymalizuje spersonalizowany wybór filmów dla poszczególnych użytkowników. Tym samym, system zbiera ogromne ilości danych na temat naszych zainteresowań i zwyczajów, które reprezentować można za pomocą macierzy.

Poniżej przedstawiamy prosty przykład zastosowania elementarnych działań na macierzach do konstrukcji algorytmu rekomendacji w systemach typu Netflix. Załóżmy, że w serwisie mamy trzech użytkowników A, B, C oraz cztery filmy, które oznaczamy przez I, II, III, IV. Dla uproszczenia modelu zakładamy również, że mamy tylko dwa gatunki filmowe, do których przyporządkowane są te cztery filmy. Zebrane dane dotyczące preferencji gatunków filmowych poszczególnych użytkowników zapisujemy w macierzy, której wierszami są użytkownicy, a kolumnami poszczególne gatunki filmowe. Druga macierz, to macierz klasyfikacji wybranych filmów do poszczególnych gatunków. Model ten przedstawiono na rysunku 2.

	Film I	*Film II*	*Film III*	*Film IV*
Gatunek 1	0,2	0,4	0,1	0,7
Gatunek 2	0,8	0,6	0,9	0,3

	Gatunek 1	*Gatunek 2*
Osoba A	0,2	0,8
Osoba B	0,3	0,7
Osoba C	0,1	0,9

Osoba A	0,68	0,56	0,74	0,38
Osoba B	0,62	0,54	0,66	0,42
Osoba C	0,74	0,58	0,82	0,34

Rys. 2. Przykładowy schemat działania algorytmu rekomendacji

Następnie wykonując elementarne operacje na macierzach, a dokładniej, mnożąc te dwie macierze, jesteśmy w stanie podać z jakim prawdopodobieństwem dany film spodoba się wybranemu użytkownikowi. Analizując nowo utworzoną macierz, wnioskujemy, że wszystkie osoby A, B i C najchętniej obejrzą film III, ponieważ w komórce trzeciej (wiersz odpowiednio A, B, C i kolumna III) mamy największe prawdopodobieństwo. W ten sposób serwisy internetowe analizując zebrane dane rozpoznają preferencje filmowe użytkowników i typują kolejne filmy w sposób zindywidualizowany.

Należy podkreślić, że opisany powyżej schemat jest bardzo uproszczony pozwalając na bardziej intuicyjną interpretację danych, ale dzięki temu może stanowić przykładowe zadanie ćwiczeniowe dla studentów na zajęciach z algebry liniowej.

W rzeczywistości serwisy takie jak Netflix czy YouTube mają miliony użytkowników i tysiące filmów do wyboru. Ponadto, preferencje użytkowników zmieniają się w czasie i stąd zachodzi potrzeba nieustannego przetwarzania i analizy ogromnych zbiorów danych.

3.2. Regresja wieloraka

Jednym z prostszych algorytmów uczenia maszynowego jest regresja wieloraka badająca zależność zmiennej objaśnianej od co najmniej dwóch zmiennych objaśniających, [5]. Przypuśćmy, że mamy dane dotyczące pewnej liczby (m) transakcji sprzedaży mieszkań, których cena (y) zależy od liczby pokoi (x_1), powierzchni mieszkania (x_2) oraz wieku mieszkania (x_3). Prosty model przewidujący cenę mieszkania w kolejnych transakcjach możemy zapisać za pomocą równania

$$\mathrm{y} = \mathrm{h}(\mathrm{x}) = \mathrm{b}_0 + \mathrm{b}_1\mathrm{x}_1 + \mathrm{b}_2\mathrm{x}_2 + \mathrm{b}_3\mathrm{x}_3.$$

Algorytm regresji wielorakiej wyznacza wartości współczynników $\mathrm{b}_0, \mathrm{b}_1, \mathrm{b}_2, \mathrm{b}_3$ minimalizujących funkcję kosztu

$$\mathrm{J}(\boldsymbol{b}) = \frac{1}{2\mathrm{m}}\sum_{\mathrm{i}=1}^{\mathrm{m}}\left[\mathrm{h}\left(\mathrm{x}^{(\mathrm{i})}\right) - \mathrm{y}^{(\mathrm{i})}\right]^2,$$

gdzie $\mathrm{x}^{(\mathrm{i})} = \left(\mathrm{x}_1^{(\mathrm{i})}, \dots, \mathrm{x}_3^{(\mathrm{i})}\right)$, $\mathrm{y}^{(\mathrm{i})}$ to odpowiednio parametry i cena mieszkania i-tej transakcji, zaś $\boldsymbol{b} = (\mathrm{b}_0, \dots, \mathrm{b}_3)$. Zauważmy, że skoro funkcja $\mathrm{J}(\boldsymbol{b})$ jest wypukła, możemy wyznaczyć ciąg punktów $b^{(\mathrm{k})}$, $\mathrm{k} = 1, \dots, \mathrm{n}$, taki że punkt $\mathrm{b}^{(\mathrm{n})}$ jest dowolnie blisko szukanego minimum funkcji kosztu. W tym celu obliczamy gradient funkcji J w punkcie $b^{(\mathrm{k})}$, tj.

$$\nabla\mathrm{J}\left(\mathrm{b}^{(\mathrm{k})}\right) = \left[\frac{\partial\mathrm{J}\left(\mathrm{b}^{(\mathrm{k})}\right)}{\partial\mathrm{b}_0}, \frac{\partial\mathrm{J}\left(\mathrm{b}^{(\mathrm{k})}\right)}{\partial\mathrm{b}_1}, \frac{\partial\mathrm{J}\left(\mathrm{b}^{(\mathrm{k})}\right)}{\partial\mathrm{b}_2}, \frac{\partial\mathrm{J}\left(\mathrm{b}^{(\mathrm{k})}\right)}{\partial\mathrm{b}_3}\right]$$

i przechodzimy od punktu $\mathrm{b}^{(\mathrm{k})}$ do $\mathrm{b}^{(\mathrm{k}+1)}$w kierunku wektora gradientu zgodnie ze wzorem

$$\mathrm{b}^{(\mathrm{k}+1)} = \mathrm{b}^{(\mathrm{k})} - \alpha\,\nabla\mathrm{J}\left(\mathrm{b}^{(\mathrm{k})}\right),$$

gdzie α jest współczynnikiem długości kroku. Jako punkt startowy możemy przyjąć dowolny punkt, np. (0,0,0,0). Uwzględniając wzór na pochodną funkcji złożonej otrzymujemy

$$b_j^{(k+1)} = b_j^{(k)} - \frac{\alpha}{m}\sum_{i=1}^{m}\left[h\left(x^{(i)}\right) - y^{(i)}\right]x_j^{(i)}, \qquad j = 0,1,2,3.$$

Powyższy model możemy opisać za pomocą równania macierzowego

$$y = Xb,$$

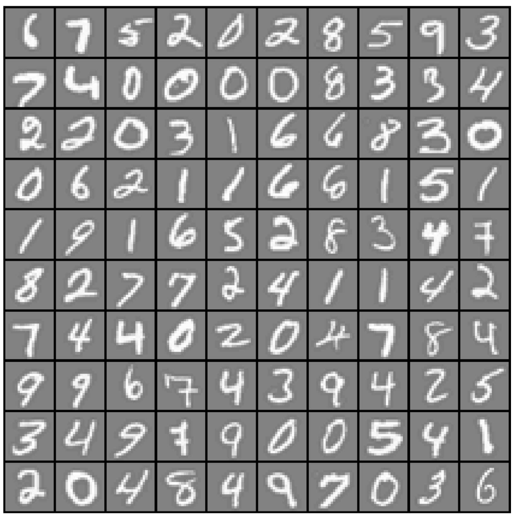
z którego uzyskujemy równanie normalne

$$b = (X^T X)^{-1} X^T y.$$

Pojawia się tutaj naturalne pytanie z zakresu algebry liniowej o istnienie macierzy odwrotnej do macierzy kwadratowej $X^T X$. W praktyce macierz ta jest na ogół macierzą odwracalną. Brak istnienia macierzy odwrotnej może świadczyć o zależności liniowej zmiennych objaśniających lub zbyt małej liczbie danych treningowych $x^{(i)}$ w stosunku do liczby zmiennych objaśniających. Wówczas do rozwiązania równania macierzowego stosujemy macierz pseudoodwrotną Moora-Penrosa. Podkreślmy, że przy dużej liczbie danych treningowych podejście algebraiczne (szukanie macierzy odwrotnej lub pseudoodwrotnej) jest dużo kosztowniejsze obliczeniowo niż podejście z zastosowaniem wektora gradientu.

4. PROBLEMY KLASYFIKACYJNE

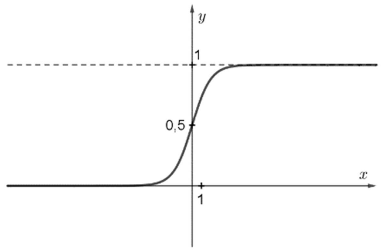
Kolejnym problemem rozwiązywanym przy pomocy uczenia maszynowego jest rozpoznawanie pisma odręcznego, należące do problemów klasyfikacyjnych. W latach 90. XX wieku automatyczne rozpoznawanie cyfr zostało zastosowane przez pocztę Stanów Zjednoczonych Ameryki Północnej do maszynowego sortowania przesyłek. Na rysunku 3. przedstawiono fragment przykładowego zestawu danych służącego do wytrenowania algorytmów. Zestaw ten składa się z bitmap 20×20 pikseli.



Rys. 3. Fragment zestawu treningowego

Zacznijmy od problemu klasyfikacji na dwie grupy. W rozwiązaniu problemu wykorzystuje się funkcję sigmoidalną (zwaną logistyczną), rysunek 4, postaci

$$g(z) = \frac{1}{1+e^{-z}}.$$

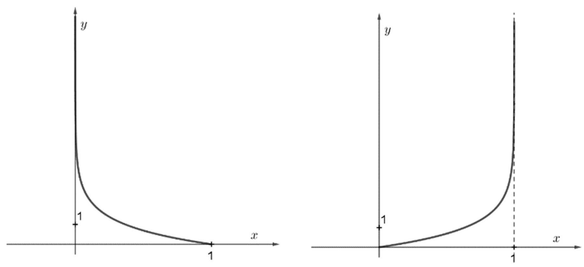


Rys. 4. Wykres funkcji sigmoidalnej

Zbiorem wartości funkcji sigmoidalnej jest przedział $(0,1)$. Załóżmy, że chcemy zakwalifikować obiekty mające cechy $x = (x_1, x_2, \dots, x_n)$ do jednej z dwóch klas. Klasom tym przyporządkowujemy etykiety 0 oraz 1. Wówczas hipotezę przynależenia obiektu do jednej z klas możemy opisać funkcją

$$h(x) = g(b_0 + b_1 x_1 + b_2 x_2 + \dots + b_n x_n),$$

gdzie $g(z)$ jest funkcją sigmoidalną. Wartość funkcji $h(x)$ interpretujemy jako prawdopodobieństwo przynależenia obiektu do klasy o etykiecie 1. Wówczas $1 - h(x)$ jest prawdopodobieństwem przynależenia obiektu do klasy o etykiecie 0. Zakładamy, że obiekt należy do klasy 1, gdy wartość $h(x) \geq 0{,}5$, a do klasy 0, gdy $h(x) < 0{,}5$. Przyjrzyjmy się wykresom dwóch funkcji logarytmicznych $y = -\log(x)$ oraz $y = -\log(1-x)$ dla argumentów w przedziale $(0,1)$, rysunek 5.



Rys. 5. Wykres funkcji $y = -\log(x)$ (po lewej) oraz $y = -\log(1-x)$ (po prawej)

Pierwsza z nich ma wartości bliskie 0 gdy argument jest bliski 1, natomiast dąży do nieskończoności, gdy argument zmierza do 0. Druga na odwrót. Własności te wykorzystuje się do konstrukcji funkcji kosztu służącej do wyznaczenia parametrów strukturalnych. Funkcja kosztu dla m danych treningowych przybiera postać

$$J(b) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left(h\left(x^{(i)}\right)\right) + \left(1 - y^{(i)}\right)\log\left(1 - h\left(x^{(i)}\right)\right)\right].$$

Wartości funkcji przynależenia do odpowiedniej klasy dla danych treningowych w zapisie macierzowym mają postać

$$h = g(Xb).$$

Zatem funkcję kosztu zapisujemy jako

$$J(b) = -\frac{1}{m}[y^T \log(h) + (1 - y^T)\log(1 - h)].$$

Dalej wykorzystując metodę gradientową otrzymujemy analogiczne, jak w przypadku regresji wielorakiej, równanie na kolejne punkty w ciągu przybliżającym punkt minimalny

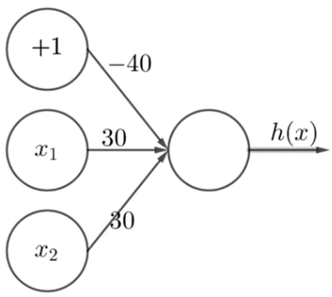
$$b_j^{(k+1)} = b_j^{(k)} - \frac{\alpha}{m}\sum_{i=1}^{m}\left[h\left(x^{(i)}\right) - y^{(i)}\right]x_j^{(i)}, \qquad j = 0,1,\dots,n.$$

Równanie to w zapisie wektorowym ma postać

$$b^{(k+1)} = b^{(k)} - \frac{\alpha}{m}X^T(g(X\boldsymbol{b}) - y).$$

W przypadku problemu klasyfikacji dla liczby klas $L \geq 3$ stosujemy strategię jeden przeciw wszystkim znajdując funkcje klasyfikujące $h_j(x)$, $j = 1,\dots,L$ dla każdej z klas. Przynależność obiektu do klasy możemy ustalić na podstawie maksymalnej wartości wyznaczonych funkcji dla klasyfikowanego obiektu, tj. $\max\left(h_1(x), h_2(x), \dots, h_L(x)\right)$.

Problemy klasyfikacyjne rozwiązuje się również stosując sieci neuronowe. Na zajęciach z matematyki dla studentów pierwszego roku, możemy zaprezentować modele neuronu wyznaczające wartość operatorów logicznych koniunkcji i alternatywy. Posługujemy się wówczas analogicznym aparatem matematycznym. Model neuronu i sieci neuronowej możemy przedstawić za pomocą odpowiedniego grafu skierowanego (rysunek 6).



Rys. 6. Przykładowy model neuronu

Możemy na przykład przyjąć funkcję

$$h(x_1, x_2) = g(-40 + 30x_1 + 30x_2) = \frac{1}{1 + e^{-(-40+30x_1+30x_2)}}.$$

W tabeli 1 zamieszczone zostały wartości funkcji koniunkcji oraz wartości uzyskane za pomocą neuronu.

Tablica 1. Wartości koniunkcji

x_1	x_2	$x_1 \wedge x_2$	$h(x_1, x_2)$
0	0	0	$h(0,0) = g(-40) \approx 0$
0	1	0	$h(0,1) = g(-10) \approx 0$
1	0	0	$h(1,0) = g(-10) \approx 0$
1	1	1	$h(1,1) = g(20) \approx 1$

Modele z kilkoma warstwami neuronów wymagają użycia algorytmów propagacji wstecznej, których przedstawienie na wspomnianych zajęciach jest zbyt trudne pojęciowo.

5. WNIOSKI KOŃCOWE

W prezentowanym artykule przedstawiliśmy propozycje wykorzystania wybranych algorytmów i metod uczenia maszynowego w motywowaniu studentów do zrozumienia potrzeby uczenia się matematyki. Uczenie maszynowe ma zastosowanie między innymi w analizie danych o sprzedaży, w wykrywaniu oszustw w transakcjach finansowych, rekomendowaniu produktów – personalizacji klienta, podejmowaniu decyzji, dynamicznym ustalaniu cen na podstawie popytu i podaży oraz rozpoznawaniu mowy.

W uczeniu maszynowym istotne znaczenie mają wybrane elementy matematyki, takie jak: działania na macierzach, składanie funkcji, obliczanie pochodnych. W związku z tym uczenie maszynowe może stanowić ciekawe uzupełnienie wykładów z algebry liniowej i analizy matematycznej pokazujące zastosowania matematyki w otaczającej nas rzeczywistości.

6. BIBLIOGRAFIA

1. Bartłomiejczyk A., Wata M.: Analizy epidemiologiczne w środowisku MATLAB, Zeszyty Naukowe Wydziału Elektrotechniki i Automatyki PG, ISSN 2353-1290, Nr 65/2019, s. 11-16.
2. Politechnika Gdańska, https://pg.edu.pl/aktualnosci/2020-10/skuteczniejsza-walka-z-guzami-nerek-dzieki-naukowcom-i-studentom-pg, (5.11.2020).
3. Deisenroth M. P., Faisal A. A., Ong Ch. Soon: Mathematics for Machine Learning, Cambridge University Press, 2020.
4. Netflix, https://help.netflix.com/pl/node/100639, (5.11.2020).
5. Kwiatkowski J., Wata M. Żuchowska M. Galikowska S.: Matematyczne metody określania wag cech rynkowych, Wycena, Wartość, Obrót, Zarządzanie Nieruchomościami, 2015, 1(110), s. 11-20.

ELEMENTS OF MACHINE LEARNING IN MATHEMATICS

The article discusses the relationship between academic mathematics course and selected topics of machine learning. It shows how simple matrix operations help VOD services to recommend film titles consistent with the interests of users, based on their previous choices. A simplified version of the multiple regression algorithm used for real estate valuation is also presented and the use of neural networks in classification problems is mentioned.

Keywords: machine learning, multiple regression, classification problems.