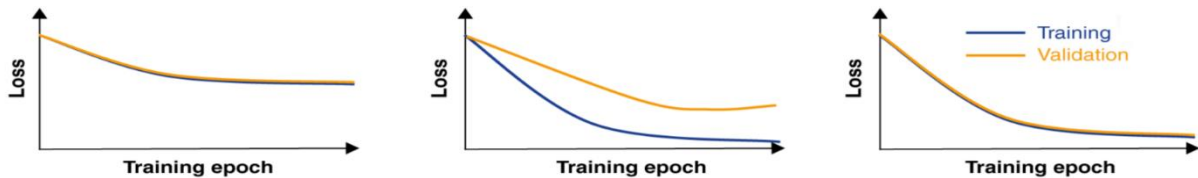


OGÓLNE

Zad1. Podaj wzory na standaryzację oraz normalizację danych. W jakim celu są one stosowane?

Zad2. Wyjaśnij znaczenie zbioru **treningowego**, **testowego** oraz **walidacyjnego**. Skomentuj poniższe krzywe uczenia.



Zad3. Czym jest **funkcja kosztu**, jaką pełni rolę procesie uczenia, w jaki sposób jest optymalizowana (podaj 3 przykładowe algorytmy), podaj 3 przykłady funkcji kosztu wraz z odpowiednimi problemami do których jest wykorzystywana.

Zad4. Jaka jest różnica pomiędzy **parametrami modelu** a **hiperparametrami**?

Zad5. Na czym polega walidacja k-krotna (**k-fold cross validation**).

Zad6. Czym różnią się **metody nadzorowane** od **nienadzorowanych** i podaj po 2 przykłady dla każdej z grup.

Zad7. Czym różnią się **modele dyskryminatywne** od **generatywnych** i podaj po 1 przykładzie metod dla każdej z grup.

Zad8. W jaki sposób reprezentowany jest obraz i dźwięk? Podaj przykłady reprezentacji dla tekstu.

KLASYCZNE METODY NAUCZANIA MASZYNOWEGO

Zad9. Podaj 3 przykładowe rozwiązania/kryteria/metody które pomagają wybrać optymalną liczbę grup w algorytmie **k-means**.

Zad10. Podaj znaczenie przekształcenia **kernel** w modelach nauczania maszynowego. Omów pokrótce jego zastosowanie do **SVM** i/lub **PCA**.

Zad11. Oceń prawdziwość poniższych stwierdzeń (P/F).

1. Dane wejściowe posiadają dwie cechy – f_1, f_2 . Stosujemy do nich algorytm **PCA** i pozostawiamy obie komponenty – $PC1, PC2$. Otrzymane komponenty będą w tym przypadku tożsame z początkowymi cechami.
2. Dane wejściowe posiadają dwie cechy – f_1, f_2 . Stosujemy do nich algorytm **PCA** i pozostawiamy obie komponenty – $PC1, PC2$. Przekształcone dane będą mieć ten sam kształt co oryginalne.
3. W algorytmie **ICA**, wyznaczanie składowych niezależnych zawsze prowadzi do jednego i optymalnego rozwiązania.
4. **ICA**, podobnie jak **PCA** działa najlepiej na danych z rozkładu normalnego.
5. Algorytmy grupowania zazwyczaj lepiej działają w przestrzeniach wysokowymiarowych (wielu zmiennych).

SIECI NEURONOWE

Zad12: Podaj znaczenie **funkcji aktywacji** w modelach neuronowych, podaj 3 przykłady takich funkcji.

Zad13: Wyjaśnij znaczenie **regularyzacji** w sieciach neuronowych. Podaj 3 przykładowe rozwiązania.

Zad14. Zbudowano sieć neuronową, która na podstawie 10 cech wejściowych przewiduje do której z dwóch kategorii należy dana obserwacja (czyli rozwiązuje problem klasyfikacji).

```
class Przykładowa_siec(nn.Module):
    def __init__(self):
        super(Przykładowa_siec, self).__init__()
        self.l1 = nn.Linear(10, 20)
        self.l2 = nn.Linear(10, 20)
        self.l3 = nn.Linear(20, 4)
        self.l4 = nn.Linear(4, 2)
    def forward(self, input):
        o1 = F.relu(self.l1(input)) + 3*F.sigmoid(self.l2(input))
        o2 = self.l3(o1)
        o3 = self.l4(o2)
        return o3
```

Dlaczego taka topologia sieci nie jest najlepszym pomysłem?

Zad15. Podaj 2 przykłady modeli sieci neuronowych opartych o **architekturę enkoder/dekoder**.

ZASTOSOWANIA

Zad16: Podaj 3 przykłady zagadnień/problemów do których można zastosować **konwolucyjne sieci neuronowe**.

Zad17. Podaj 3 przykłady zagadnień/problemów do których można zastosować **rekurencyjne sieci neuronowe**.

Zad18. Zaproponuj przykładowe 2 metody dla każdego z następujących zagadnień

- Chcemy pogrupować streszczenia artykułów naukowych na artykuły biologiczne, fizyczne i chemiczne.

- Chcemy zbudować klasyfikator przewidujący w oparciu o pewne informacje dotyczące stylu życia (jak rodzaj diety, status palenia papierosów) czy u danej osoby rozwinie się nowotwór płuc czy też nie.

- Zebraliśmy 1000 cech dotyczących mikrobiomu, dokładniej rozważano 300 gatunków bakterii sprzyjających i niebezpiecznych człowiekowi. Chcemy wytypować które z cech rozróżniają te dwie grupy.

Autoenkoder – do redukcji danych (embeddingów), można używać np. do generowania obrazów
Seq2Seq

-> może dać jeszcze pytanie w jaki sposób reprezentowany jest obraz, dźwięk, jak można reprezentować tekst

Zad2: (underfitting, overfitting, ok)

Zad4: – te drugie są ustawiane, a te pierwsze uczone podczas uczenia

Zad8: → wspomnieć o wirtualizacji obiektów i wyodrębnieniu pewnych cech, że to nigdy nie jest 1 do 1

Może jeszcze pytać o feature selection i feature extraction