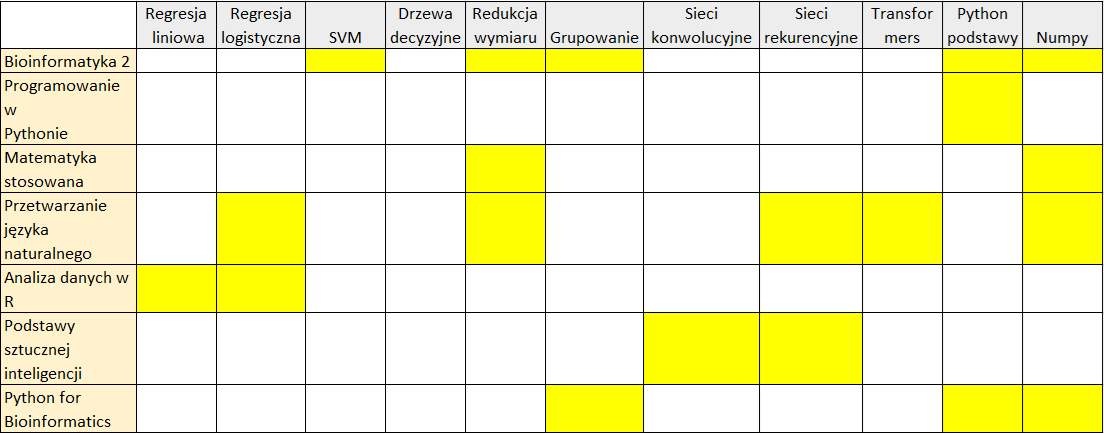
**

*Cel ćwiczeń?*

- podsumowanie i uporządkowanie metod do analizy danych

- wskazanie wad i zalet wybranych metod (w tym możliwość dostrojenia modeli za pomocą hiperparametrów)

- rozszerzenie metod wybranego typu (np. ICA w redukcji wymiaru czy algorytm Gustafsona-Kessela w grupowaniu)

- krytyczna ocena modeli

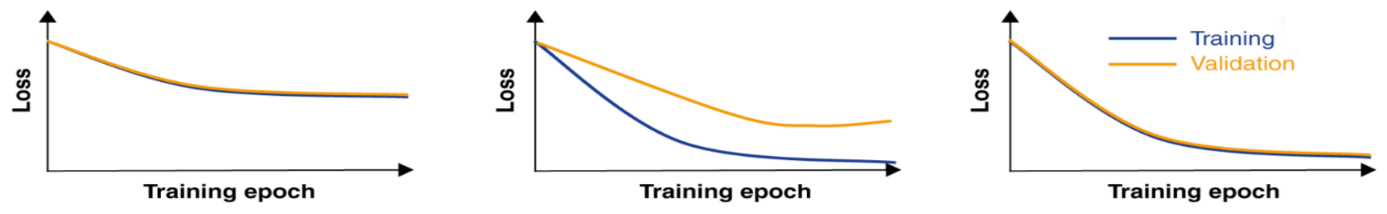
- zadania różnego typu: generowanie wyników, analiza wyników, porównanie metod, implementacja metod 🡪 zadania

- umiejętność planowania analizy danych i jej realizacja 🡪 prezentacja i projekt

*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. Podaj 3 przykładowe **algorytmy optymalizacyjne**.

Zad5. Jaka jest różnica pomiędzy **parametrami modelu** a **hiperparametrami**?

Zad6. Na czym polega walidacja k-krotna (**k-fold cross validation**).

Zad7. Czym różnią się **metody nadzorowane** od **nienadzorowanych** i podaj po 2 przykłady dla każdej z grup.

Zad8. Czym różnią się **modele dyskryminatywne** od **generatywnych** i podaj po 1 przykładzie metod dla każdej z grup.

Zad9. Podaj 3 przykładowe **miary oceny** zbudowanego modelu.

Zad10. W jaki sposób reprezentowany jest obraz i dźwięk? Podaj przykłady reprezentacji dla tekstu.

*KLASYCZNE METODY NAUCZANIA MASZYNOWGO*

Zad11. Które z poniższych zależności pomiędzy **y** a zmiennymi **x1 i x2** można zamodelować za pomocą **regresji liniowej.**

Zad12. Podaj 3 przykładowe rozwiązania/kryteria/metody które pomagają wybrać optymalną liczbę grup w algorytmie **k-means.**

Zad13. Podaj znaczenie przekształcenia **kernel** w modelach nauczania maszynowego. Omów pokrótce jego zastosowanie do **SVM** i/lub *PCA*.

Zad14. Jaka jest różnica między podejściem **feature selection** a **feature extraction**. Podać przykłady metod.

Zad15. Oceń prawdziwość poniższych stwierdzeń (P/F).

1. Dane wejściowe posiadają dwie cechy – *f1, f2*. 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 – *f1, f2.* 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).

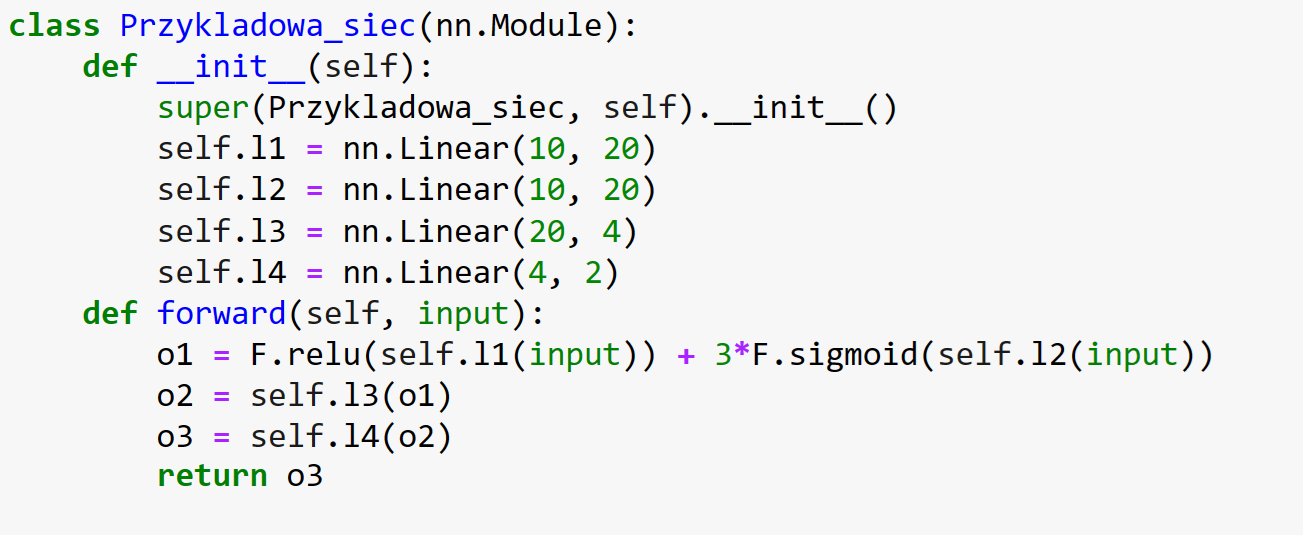
SIECI NEURONOWE

[A Neural Network Playground (tensorflow.org)](https://playground.tensorflow.org/#activation=tanh&batchSize=10&dataset=circle&regDataset=reg-plane&learningRate=0.03&regularizationRate=0&noise=0&networkShape=4,2&seed=0.57675&showTestData=false&discretize=false&percTrainData=50&x=true&y=true&xTimesY=false&xSquared=false&ySquared=false&cosX=false&sinX=false&cosY=false&sinY=false&collectStats=false&problem=classification&initZero=false&hideText=false)

Zad16: Podaj znaczenie **funkcji aktywacji** w modelach neuronowych, podaj 3 przykłady takich funkcji.

Zad17: Wyjaśnij znaczenie **regularyzacji** w sieciach neuronowych. Podaj 3 przykładowe rozwiązania.

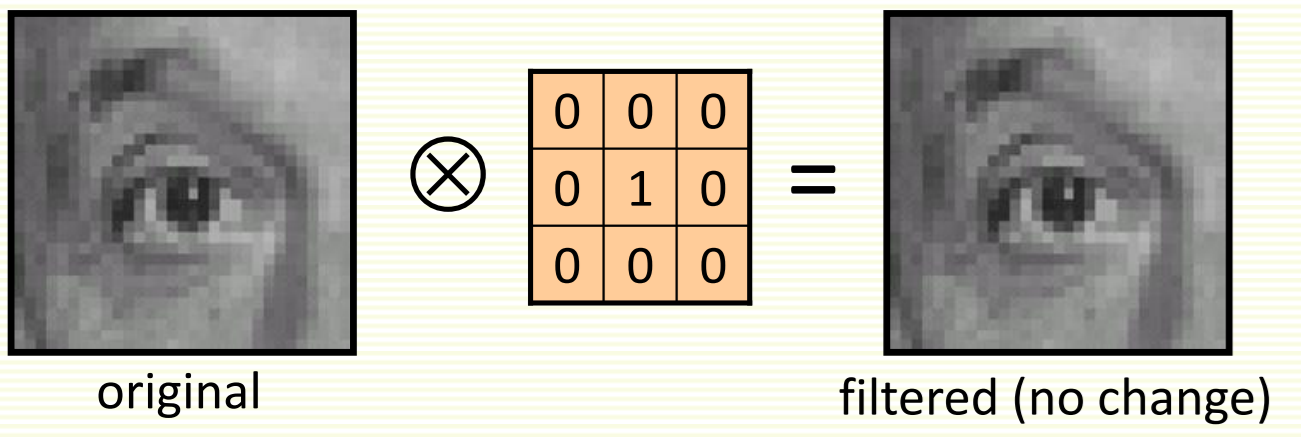
Zad18. Zbudowano sieć neuronową, która na podstawie 10 cech wejściowych przewiduje do której z dwóch kategorii należy dana obserwacja (czyli rozwiązuje problem klasyfikacji).

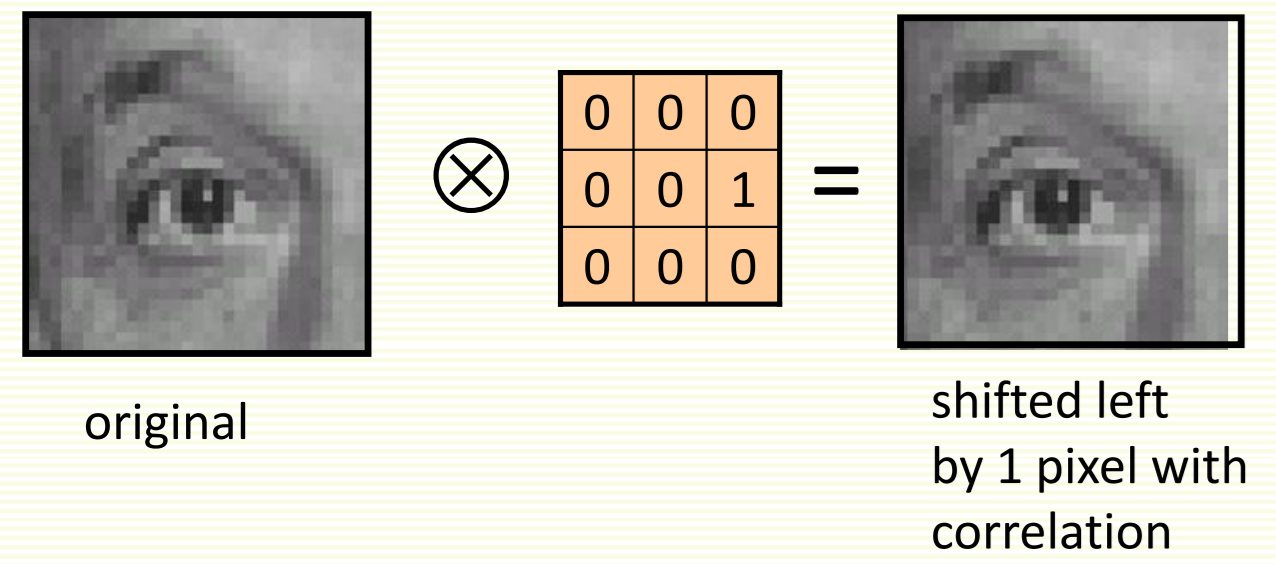


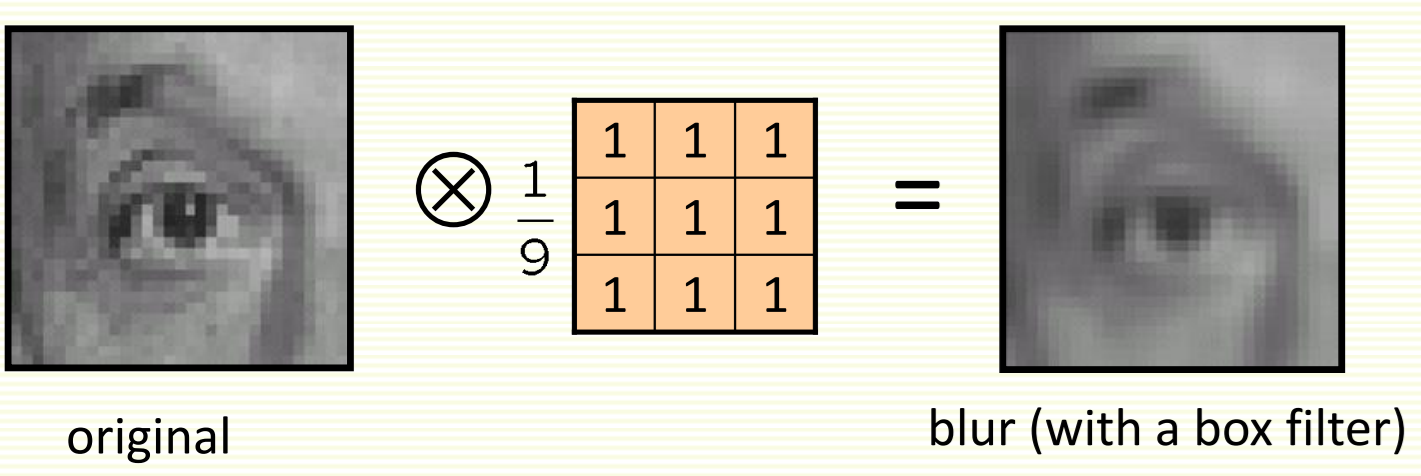
Dlaczego taka topologia sieci nie jest najlepszym pomysłem?

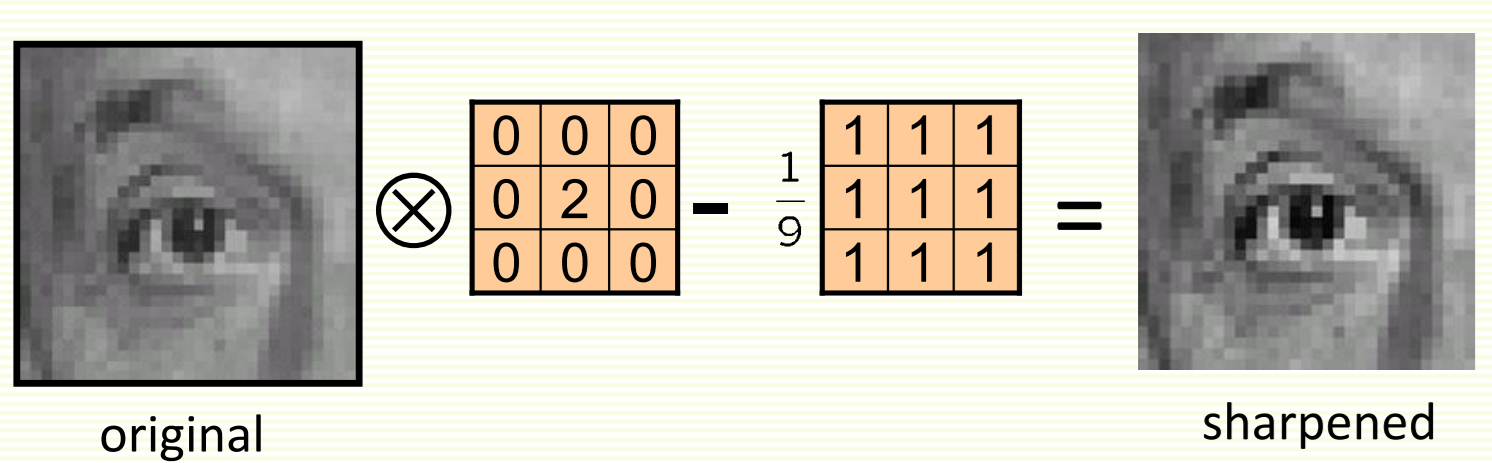
Zad19. Poniżej przedstawiono kilka filtrów do obrazów. Co robi każdy z nich?

, , , ,









Zad20. Podaj 2 przykłady modeli sieci neuronowych opartych o **architekturę enkoder/dekoder**.

Zad21. Czym jest **weight sharing**?

Zad22. Czym jest **finetuning**?

*ZASTOSOWANIA*

Zad23: Podaj 3 przykłady zagadnień/problemów do których można zastosować **konwolucyjne sieci neuronowe**.

Zad24. Podaj 3 przykłady zagadnień/problemów do których można zastosować **rekurencyjne sieci neuronowe.**

Zad25. Zaproponuj przykładowe 2 metody dla każdych z następujących zagadnień

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

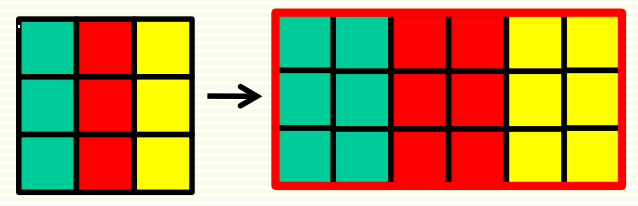
- Chcemy zbudować klasyfikator przewidujący w oparciu o pewne informacje dotyczące stylu życia (jak rodzaj diety, status palenia papierosów) czy u danej osoby rozwinie się nowotwór płuc czy też nie.

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

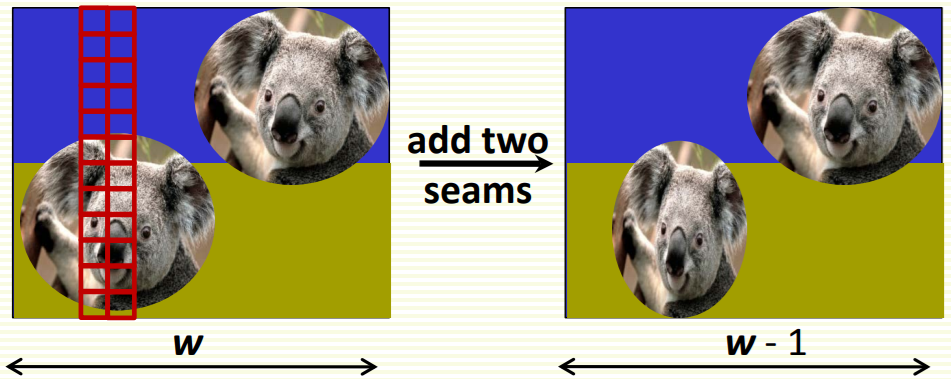
Inne

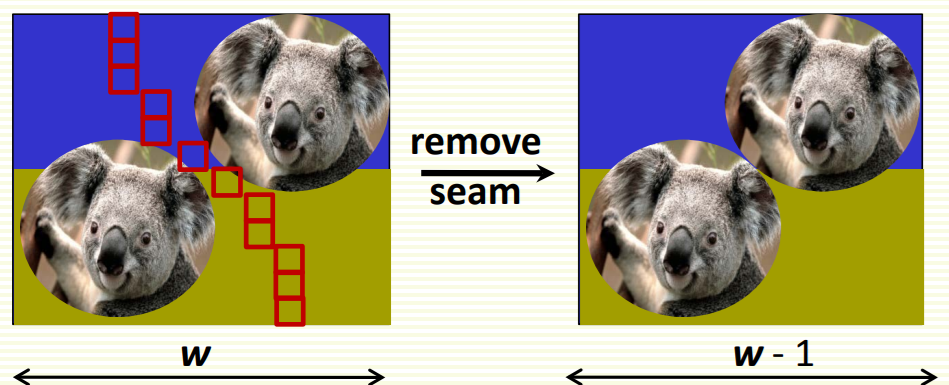
Zad26. Zmiana rozmiaru obrazka

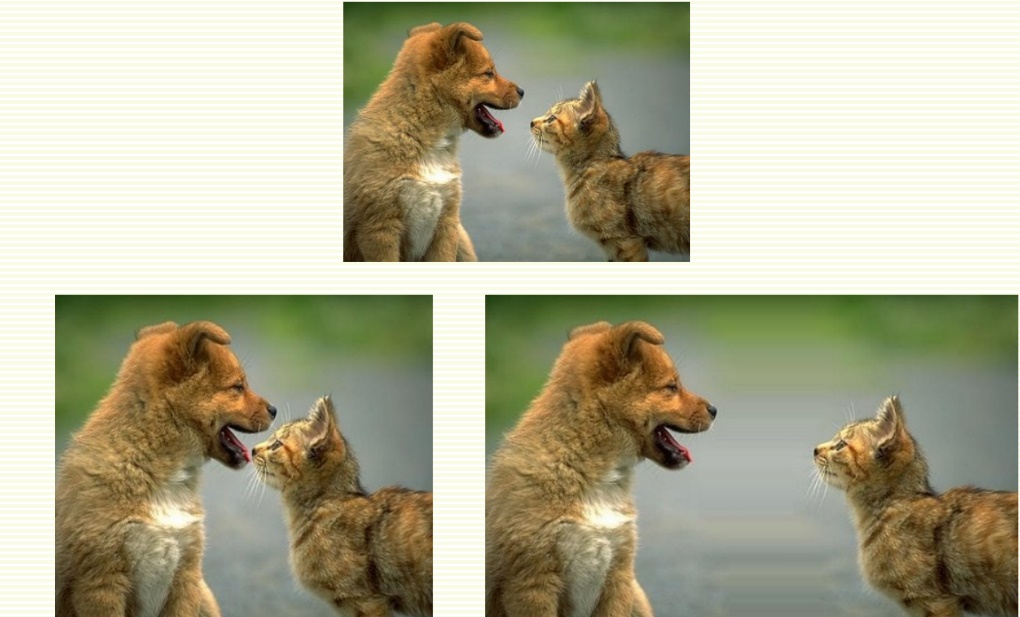


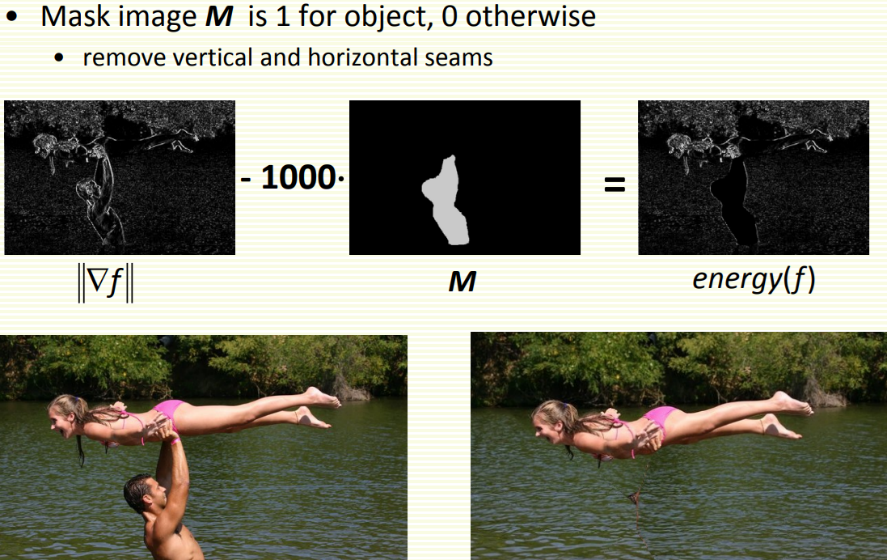


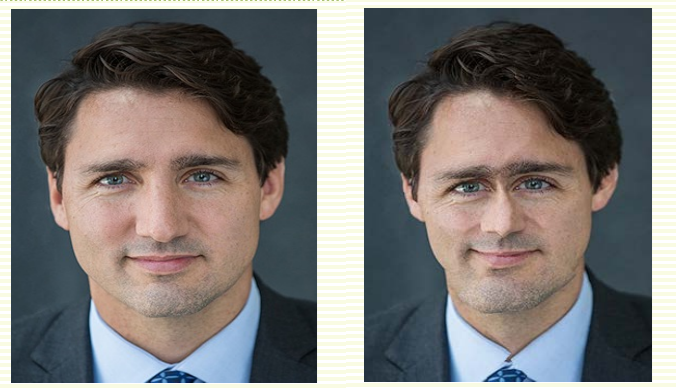












Autoenkoder – do redukcji danych (embeddingów), można uzywac np. do generowania obrazów

Seq2Seq

Zad2: (underffiting, overffitjng, ok)

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

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

Zad9: accuracy, f-score, top-5-error I inne (czulosc, specfyicznosc)

Wzad11: ostatni nie, chyba ze zalozymy np. ze zmienne są nieujemne (tak to nie jest funkcja)