

KNSI Golem Bootcamp 2021

Spotkanie 1

Agenda

- Czym jest KNSI Golem?
- Jakie mamy plany na ten (i przyszły) semestr?
- Omówienie bootcampu
- Część druga ;)

KNSI Golem, a co to takiego?

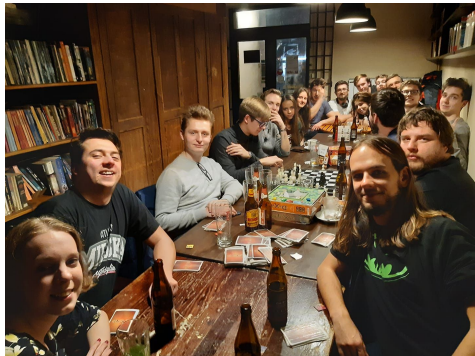
Kluczowe informacje

- Istniejemy od 2017r.
- Wydział EiTI
- opiekun koła: dr hab. Paweł Wawrzyński
- Zarząd na ten rok akademicki:
 - Antoni Kowalczuk (prezes)
 - Maciej Chrabąszcz (vice)
 - Piotr Kitłowski (skarbnik)
- Własny komputer obliczeniowy w ZSI (4xGeForce RTX 2070 Super)



Aktywność koła

Prelekcje + Piwo



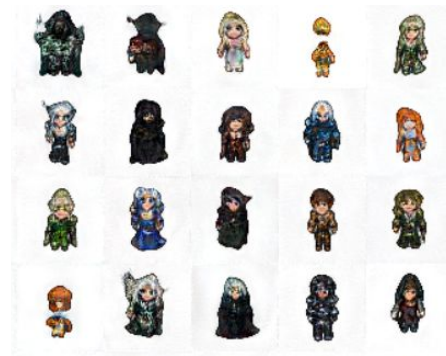
Hackathony



Bootcamp



Projekty



Roadmap, czyli jak zostać Golemem

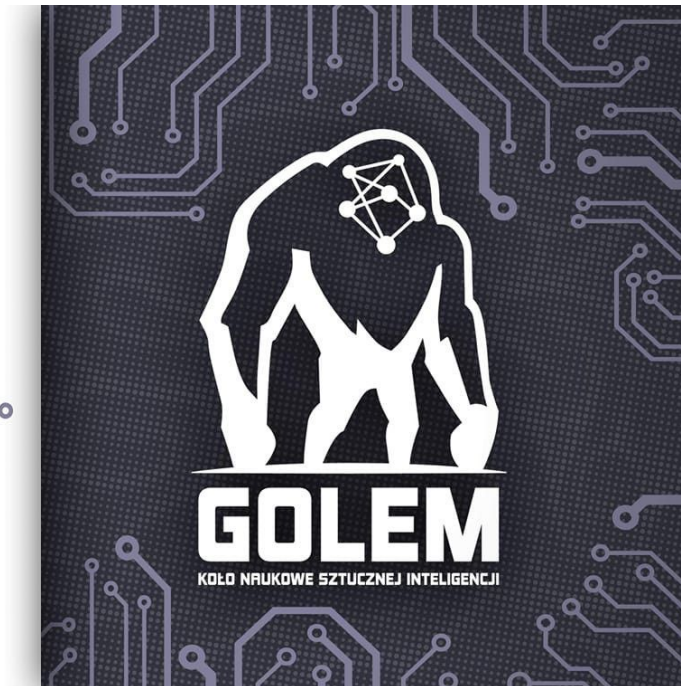
1. Bootcamp

Jeśli tu jesteś, to dopiero początek Twojej drogi :)

BOOTCAMP
Wprowadzenie

Antoni Kowalczuk
Maciej Kaczowski

21 PAŹDZIERNIKA
sala 170
WEiTI



2. Bootcamp 2: Electric Boogaloo (level-up)

- Termin TBD
- Wprowadzenie do DNN
- Rozszerzenie wiedzy zdobytej na Bootcampie
- Cel: dostarczyć wiedzę i umiejętności niezbędne w wystartowaniu w planowanym hackathonie (szczegóły TBD)
- Tak zwana ostra jazda xD

3. Pytorch Zero2Hero

- Top Framework do DNN
- Forma i terminy jeszcze TBD
- Cel: zaznajomić z podstawowymi i zaawansowanymi architekturami DNN i ich implementacją w pytorch





DLACZEGO
PO CO?

Zakład Sztucznej Inteligencji

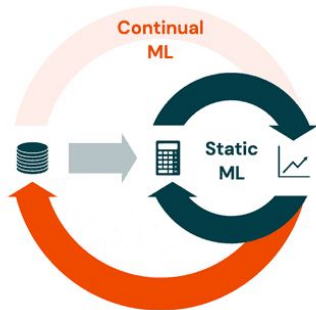
- Projekty badawcze
- Specjaliści i mentorzy
- Ciekawe i zaawansowane tematy
- Spin-off
- Co jest potrzebne:
 - skille zdobyte w poprzednich punktach
 - chęci
 - czas

**A przede wszystkim potrzebny
jest zespół**

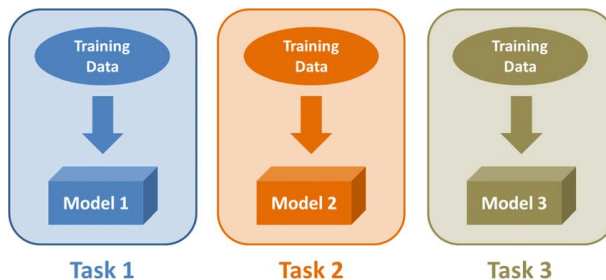


Co się dzieje w ZSI?

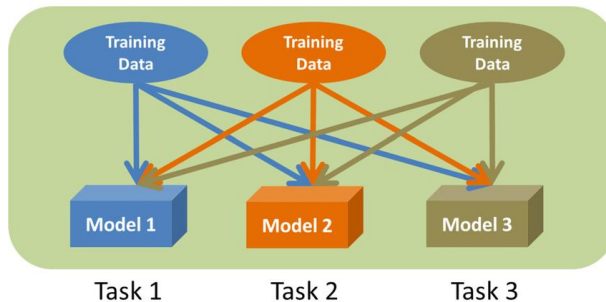
Continual Learning



Multitask Learning

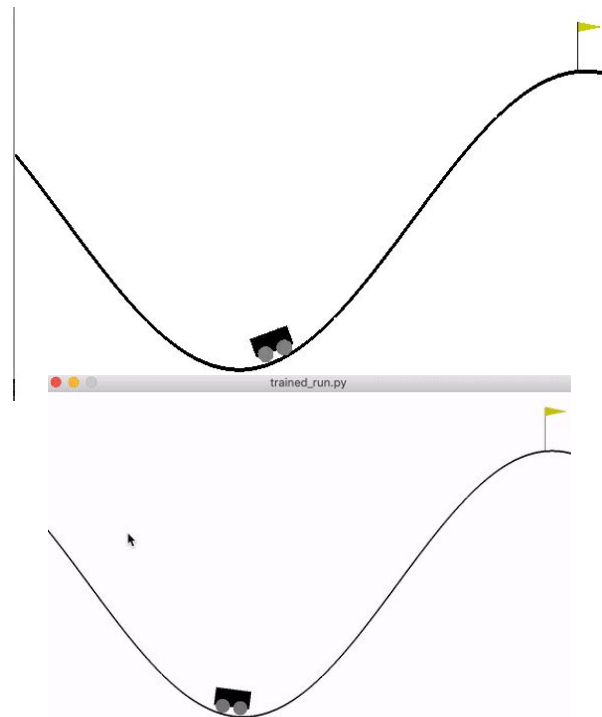


(a)



(b)

Reinforcement Learning



Bootcamp – struktura

1. 5 spotkań:

- a. Wstęp i omówienie biblioteki pandas (Antoni x Maciek)
- b. Regresja (Weronika)
- c. Klasyfikacja (znowu Antoni xD)
- d. Tips&Tricks (Anton)
- e. Sztuczne Sieci Neuronowe (Piotr)

2. W ramach każdego spotkania:

- a. Wykład, prezentacja + przykłady w kodzie
- b. Praca domowa wraz z rozwiązaniami

3. Po trzecim spotkaniu: konkurs na Kagglu (TBD)

Q&A

Część druga

Agenda po raz drugi

- Dane, dane, dane – kluczowe koncepty
- Typy zmiennych
- Jak rozumieć Machine Learning?
- Czym jest predykcja?
- Podstawowe zadania
- Metody uczenia się
- Podsumowanie
- Część trzecia ;)



O danych słów kilka

Kluczowe koncepty

- Zbiory:
 - treningowy
 - testowy
- Zbiór cech (tzw. features)
- Cel predykcji (target, label)
- Dwuwymiarowe tabelki (przynajmniej na razie)
- Typy cech:
 - ciągłe
 - dyskretne (kategoryczne)

Dataset

Cechy

Cel

Wymiar cech

Wymiar rekordów (wiersze)

	categorical_feature_1	categorical_feature_2	categorical_feature_3	categorical_feature_4	continuous_feature_1	continuous_feature_2	target
0	f	d	h	g	335.699617	145.288619	0
1	b	h	e	e	101.786116	10.796348	1
2	f	g	h	f	251.530238	205.116568	0
3	h	h	c	f	42.022427	28.089197	0
4	c	e	g	h	114.155232	12.432317	1

Machine Learning (Uczenie Maszynowe)

Jak to rozumieć?

Maszynowe

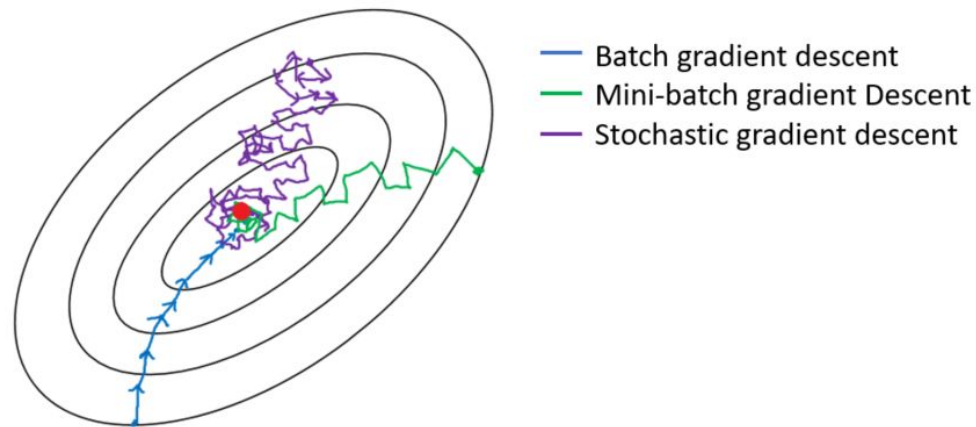
- Bez tzw. 'ifologii'
- Oparte na danych
- Celem: parametryzowany model rozwiązujący pewne, wysoce wyspecjalizowane zadanie (np. rozpoznawanie czy dany zbiór pikseli reprezentuje literę G)

Uczenie się

- Inspirowane sposobem, w jaki uczy się człowiek
- Mechanizm: pętla sprzężenia zwrotnego
- Kluczem: dane

Model

Algorytm uczący



Parametry

$$\begin{matrix} & \begin{matrix} 1 & 2 & \dots & n \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ \vdots \\ m \end{matrix} & \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ a_{31} & a_{32} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \end{matrix}$$

Funkcja straty

$$\text{MSE}(\mathcal{D}, f) := \frac{1}{n} \sum_{i=1}^n \|\mathbf{y}_i - f(\mathbf{x}_i)\|_2^2,$$

Trening

Zbiór treningowy

	categorical_feature_1	categorical_feature_2	categorical_feature_3	categorical_feature_4	continuous_feature_1	continuous_feature_2	target
0	e	c	c	e	118.633799	152.453576	0
1	e	e	h	c	204.479752	135.259612	1
2	h	g	e	f	99.076089	25.263754	0
3	e	a	b	a	20.543345	263.686608	1
4	a	c	c	e	137.473731	190.692092	1
5	b	d	c	h	53.853686	42.772230	1
6	f	c	c	a	176.981080	148.681766	0
7	b	b	e	e	167.223956	41.521861	0
8	e	e	e	f	219.476611	276.671168	1
9	a	b	g	f	200.718175	174.592773	1
10	a	g	b	b	116.951042	66.945875	1
11	d	h	f	d	75.236034	19.340713	0
12	c	f	b	e	32.082324	169.998350	1
13	h	c	a	h	55.152452	271.143753	1
14	g	g	e	c	151.868236	91.840448	0
15	b	g	f	e	78.591005	305.604886	0
16	g	a	c	g	159.383117	236.990436	1
17	h	b	g	d	33.673526	164.716809	1
18	a	e	g	d	200.276430	325.659930	0
19	c	e	b	c	211.581914	175.281508	0

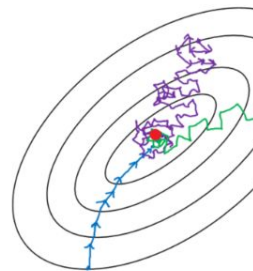
Aktualizacja parametrów modelu

$$\begin{matrix}
 & \begin{matrix} 1 & 2 & \dots & n \end{matrix} \\
 \begin{matrix} 1 \\ 2 \\ 3 \\ \vdots \\ m \end{matrix} & \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ a_{31} & a_{32} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}
 \end{matrix}$$

Predykcja

$$\text{MSE}(\mathcal{D}, f) := \frac{1}{n} \sum_{i=1}^n \|y_i - f(\mathbf{x}_i)\|_2^2,$$

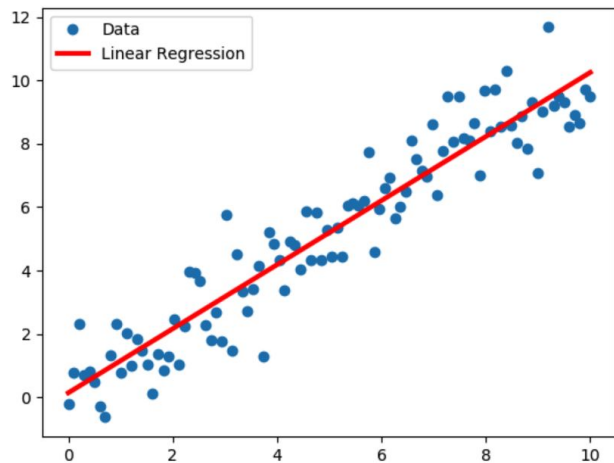
Błąd / Strata



Typy zadań predykcyjnych (podstawowe)

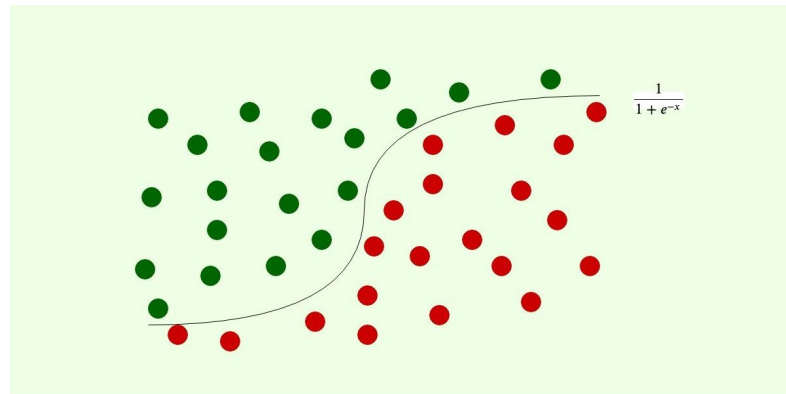
Regresja (spotkanie 2.)

- Przewidywanie ciągłej wartości
- Przykład: jaka powinna być wartość nieruchomości bazując na danych o niej?
- Podstawowy model: regresja liniowa



Klasyfikacja (spotkanie 3.)

- Przewidywanie klasy/kategorii
- Przykład: jaki gatunek piwa powstanie mając dane nt procesu jego wytwarzania?
- Podstawowy model: drzewo decyzyjne, regresja logistyczna (tak, to coś innego)



Typy uczenia się

Nadzorowane (supervised)

- W zbiorze treningowym znana jest nam wartość, którą chcemy przewidywać
- Naszym zadaniem jest przewidzenie jej w zbiorze testowym (ewaluacyjnym)
- Bootcamp dotyczy praktycznie tylko takiego typu uczenia się

Nienadzorowane (unsupervised)

- Zazwyczaj dotyczy klastryzacji
- Przykład: mając 100 punktów danych o zachowaniu użytkowników w serwisie znajdź ukryte podgrupy
- (Jeśli działa) umożliwia odkrywanie ukrytych zależności w dużych i wielowymiarowych zbiorach danych

Podsumowanie

- KNSI Golem jest fajne i robi fajne rzeczy, róbcie je z nami :)
- Od zera do bohatera – roadmap
- Struktura bootcampu
- Dane
- Kluczowe koncepty: Maszynowe Uczenie się
- Z czego składa się model ML?
- Jak wygląda jego trening?
- Typy zadań oraz uczenia się

Q&A

Dzięki, zapraszamy na część trzecią :)