

Komitety sieci konwolucyjnych w zagadnieniu
klasyfikacji przy jednoczesnym zaszumieniu
danych wejściowych oraz etykiet klas

Stanisław Kaźmierczak

Agenda

- Szum i jego rodzaje
- Źródła szumu
- Model
- Architektura sieci
- Zaszumianie danych
- Wyniki
- Analiza błędów
- Propozycje dalszych prac



Wprowadzenie

- Problem klasyfikacji obrazów jest obecnie bardzo popularnym zagadnieniem
- Modele trenowane na dobrej jakości danych osiągają w niektórych dziedzinach lepsze rezultaty niż ludzie
- Rzeczywiste dane są dość często zdegenerowane szumem występującym na atrybutach (cechach) lub etykietach (klasach)
- Wpływa negatywnie na działanie klasyfikatorów



Szum

- Definicja
 - *Cokolwiek co zakłóca, przyciemnia relację między cechami obiektu, a jej klasą [1, 3]*
 - *Niesystematyczny błąd [2, 3]*
- Rodzaje szumu [2, 4]
 - Szum związany z atrybutami (ang. *Feature/attribute noise*)
 - Szum związany z etykietami (ang. *Label/class noise*)
- Szum na etykietach jest potencjalnie bardziej szkodliwy [4]
 - Jest wiele cech, ale tylko jedna etykieta
 - Znaczenie (istotność) każdej cechy na proces uczenia jest różna, natomiast etykiety mają zawsze duży wpływ



Źródła szumu na etykietach

- Informacja dostarczona ekspertowi jest niewystarczająca do przeprowadzenia niezawodnego etykietowania
 - Przykład: decyzja o chorobie na podstawie wywiadu lekarskiego
- Błąd eksperta (człowieka lub maszyny)
 - Etykiety dostarczone przez nie-ekspertów, np. *Amazon Mechanical Turk*
- Nieobiektywność ekspertów
 - Przykład: eksperci rzadko zgadzają się odnośnie dokładnych granic we wzorcach sygnału w analizie EKG
- Sposób kodowania lub problemy komunikacyjne
 - Przykład: błędne oznaczenie maili jako spam wynikające z nierozumienia mechanizmu lub przypadkowego kliknięcia



Źródła szumu na atrybutach

- Niedostateczna jakość urządzeń pomiarowych
- Wynikające z praw fizyki
 - Przykład: wibrujące pole elektromagnetyczne
- Warunki pomiaru
 - Przykłady: hałas w tle, słabe oświetlenie
- Konwersja sygnału analogowego na cyfrowy
- Konsekwencja kompresji danych
- ...



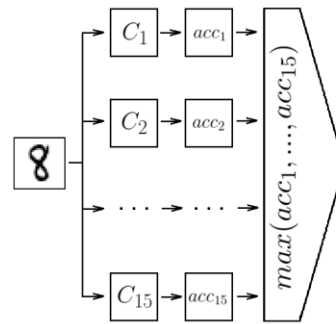
Klasyfikacja danych zaszumionych

- Wraz ze wzrostem zniekształcenia obrazów, jakość klasyfikatorów spada istotnie szybciej niż jakość klasyfikacji dokonywanej przez człowieka
- Szum na etykietach ma również wiele negatywnych konsekwencji
 - Pogorszenie jakości predykcji
 - Potrzebny bardziej złożony model
 - Potrzebny większy zbiór treningowy i/lub dłuższy proces uczenia
- Na uwadze trzeba mieć również fakt, że pozyskanie poprawnie otykietowanych danych jest z reguły czasochłonne, kosztowne i wymaga znaczącego zaangażowania ekspertów dziedzinowych
- Z przeprowadzonego przeglądu literatury wynika, że problem klasyfikacji obrazów przy jednoczesnym zaszumieniu atrybutów i etykiet nie był do tej pory analizowany

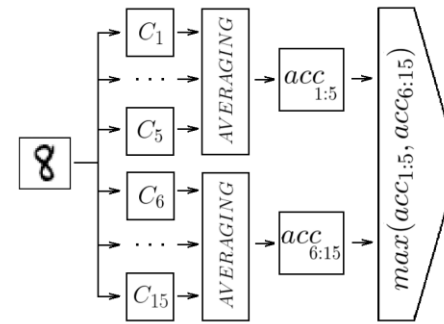
Model

- Założenia
 - Brak wiedzy na temat rodzaju i stopnia zaszumienia atrybutów oraz etykiet
 - Działanie na *surowych* obrazach (bez żadnego preprocessingu)
- CNNy, w ogólności, osiągają dobre rezultaty na zaszumionych obrazach (szum na atrybutach)
- Zespoły modeli dobrze radzą sobie z szumem na etykietach
- Jak komitet CNNów poradzi sobie z jednoczesnym szumem atrybutów i etykiet?
- Każdy z pojedynczych modeli uczony jest na całym zbiorze danych
- Dla ustalonego N (liczba pojedynczych modeli w Komitecie), wyniki N -komitetu porównywane są z wynikami zbioru mniejszych komitetów oraz zbioru pojedynczych CNNów, których suma wynosi N

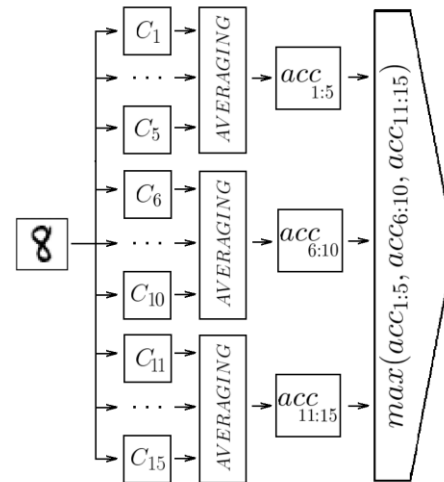
Przykład dla $N = 15$



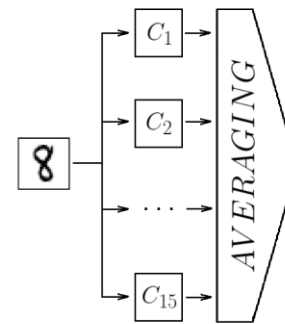
(a)



(b)



(c)



(d)

- (a) Kombinacja 15-stu pojedynczych CNNów
- (b) Kombinacja 5-komitetu i 10-komitetu
- (c) Kombinacja 3-ech 5-komitetów
- (d) 15-komitet

Zbiory danych

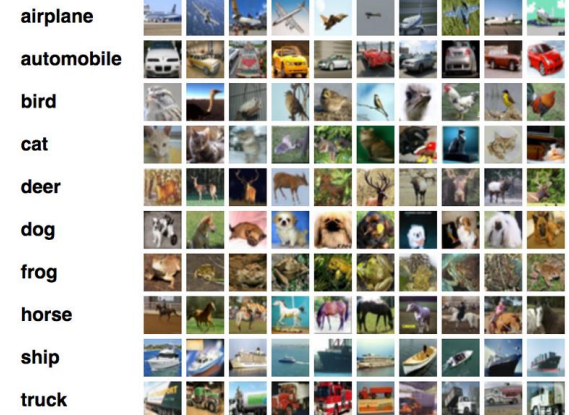
- *MNIST*

- Zbiór pisanych ręcznie cyfr (10 klas)
- Obrazy 28x28, w skali szarości
- Zbiór treningowy – 60k egz.
- Zbiór testowy – 10k egz.

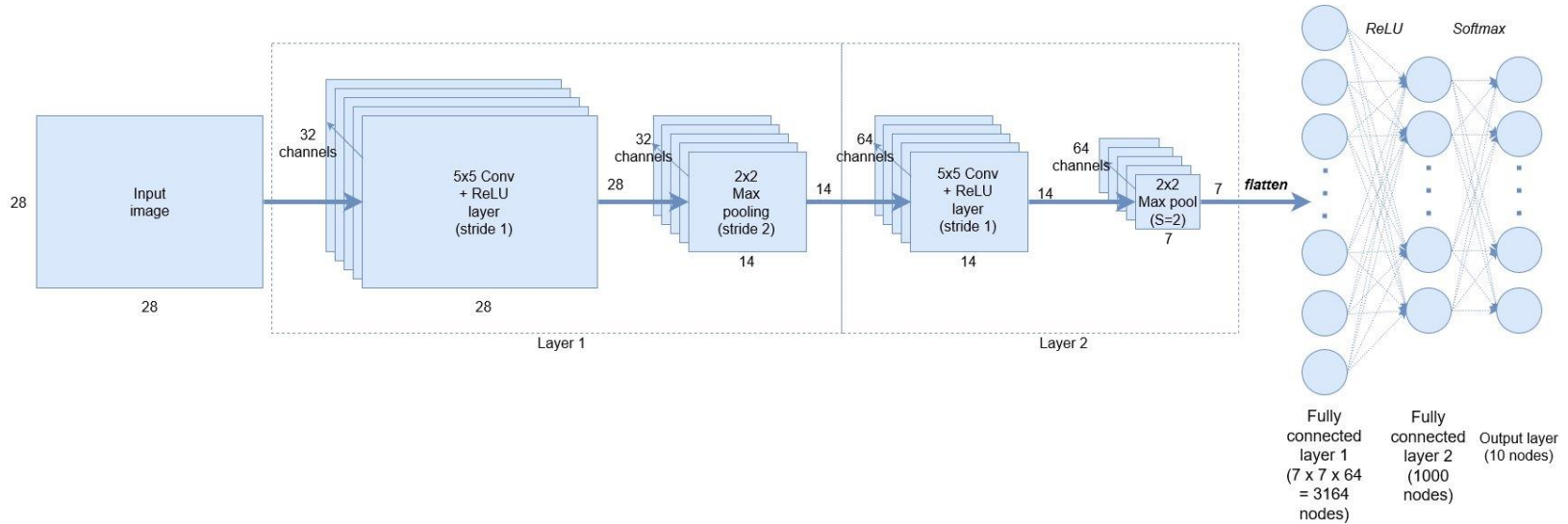


- *CIFAR-10*

- Zbiór obiektów należących do 10 klas
- Obrazy 32x32, kolorowe (3 kanały)
- Zbiór treningowy – 50k egz.
- Zbiór testowy – 10k egz.



Architektura sieci (MNIST)



[5]



Architektura sieci (CIFAR-10)

- convolutional layer, 32 output channels with a size of 3x3, ReLU activation function
- convolutional layer, 64 output channels with a size of 3x3, ReLU activation function
- max pooling layer of size 2x2
- dropout set to 25%
- convolutional layer, 128 output channels with a size of 3x3, ReLU activation function
- max pooling layer of size 2x2
- convolutional layer, 128 output channels with a size of 3x3, ReLU activation function
- max pooling layer of size 2x2
- dropout set to 25%
- flatten layer
- fully-connected layer with 1024 neurons and ReLU activation function
- dropout set to 50%
- fully-connected layer with 10 neurons and softmax activation function

Zaszumianie danych

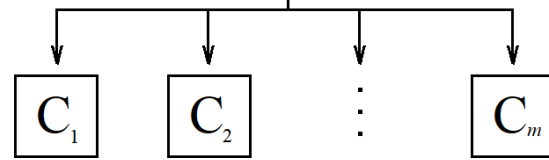
- Wartości atrybutów oryginalnego zbioru zostały przeskalowane liniowo do przedziału $[0, 1]$
- 5 kopii oryginalnego zbioru zostało zaszumionych szumem gaussowskim z odchyleniem standardowym $\sigma = 0.05, 0.15, 0.25, 0.35, 0.45$
- Wartości mniejsze od 0 zostały zastąpione przez 0
- Wartości większe od 1 zostały zastąpione przez 1
- Dla oryginalnego zbioru i każdej z 5 kopii, utworzonych zostało kolejne 5 kopii, a wartości etykiet zostały zmienione z prawdopodobieństwem $p = 0.05, 0.15, 0.25, 0.35, 0.45$
- Jeśli dana etykieta została wylosowana do zmiany, nowa wartość losowana była z rozkładu dyskretnego jednostajnego $U\{0, 9\}$, aż do uzyskania innej niż oryginalna wartości
- Uzyskujemy w ten sposób 36 wersji każdego ze zbiorów danych (MNIST i CIFAR-10)

Predykcja

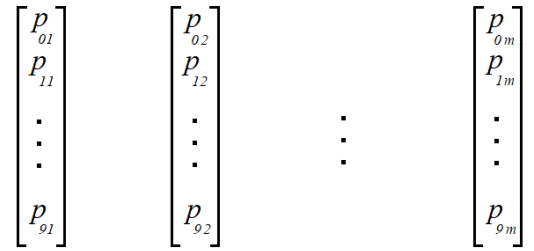
Test sample



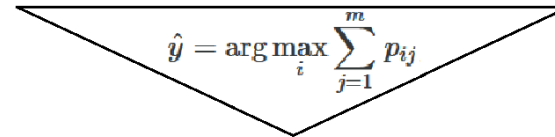
Classification models



Predicted class probabilities



Soft voting



Final prediction





Eksperyment

- 20% danych treningowych jest używanych w każdej epoce jako zbiór walidacyjny (w każdej epoce losowany jest nowy zbiór walidacyjny)
- Etykiety na zbiorze testowym nie są zaszumiane (chcemy mieć wiarygodną ocenę działania modelu)
- Dla każdego z 36 poziomów szumu zostało nauczonych niezależnie 50 CNNów
- N -komitety zostały utworzone poprzez losowanie bez zwracania N modeli z powyższej puli
- Dla każdego poziomu zaszumienia eksperyment został wykonany 30 razy

Wyniki dla MNIST

MNIST – TEST ERROR RATE [%] OF A SINGLE CNN

Feature noise	Label noise						AVG
	$p = 0$	$p = 0.05$	$p = 0.15$	$p = 0.25$	$p = 0.35$	$p = 0.45$	
-	0.83 ± 0.08	0.93 ± 0.09	1.08 ± 0.10	1.31 ± 0.14	1.44 ± 0.11	1.73 ± 0.12	1.22
$N(0, 0.05)$	0.84 ± 0.09	0.96 ± 0.08	1.17 ± 0.11	1.27 ± 0.11	1.45 ± 0.12	1.72 ± 0.15	1.24
$N(0, 0.15)$	0.95 ± 0.08	1.07 ± 0.09	1.34 ± 0.10	1.52 ± 0.11	1.73 ± 0.15	1.91 ± 0.12	1.42
$N(0, 0.25)$	1.21 ± 0.12	1.25 ± 0.09	1.54 ± 0.12	1.74 ± 0.14	1.97 ± 0.12	2.44 ± 0.20	1.69
$N(0, 0.35)$	1.55 ± 0.12	1.84 ± 0.11	2.02 ± 0.12	2.28 ± 0.19	2.77 ± 0.20	3.14 ± 0.30	2.27
$N(0, 0.45)$	2.36 ± 0.17	2.67 ± 0.16	3.06 ± 0.18	3.61 ± 0.16	4.14 ± 0.22	4.76 ± 0.35	3.43
AVG	1.29	1.45	1.70	1.96	2.25	2.62	1.88

MNIST – TEST ERROR RATE [%] OF A 5-COMMITTEE

Feature noise	Label noise						AVG
	$p = 0$	$p = 0.05$	$p = 0.15$	$p = 0.25$	$p = 0.35$	$p = 0.45$	
-	0.53 ± 0.04	0.62 ± 0.04	0.74 ± 0.05	0.96 ± 0.04	1.01 ± 0.07	1.24 ± 0.07	0.85
$N(0, 0.05)$	0.54 ± 0.04	0.62 ± 0.06	0.80 ± 0.06	0.89 ± 0.06	1.08 ± 0.04	1.21 ± 0.07	0.86
$N(0, 0.15)$	0.59 ± 0.04	0.73 ± 0.04	0.92 ± 0.07	1.06 ± 0.07	1.22 ± 0.06	1.29 ± 0.07	0.97
$N(0, 0.25)$	0.78 ± 0.05	0.85 ± 0.05	1.11 ± 0.04	1.26 ± 0.06	1.35 ± 0.07	1.68 ± 0.08	1.17
$N(0, 0.35)$	1.04 ± 0.05	1.23 ± 0.06	1.42 ± 0.05	1.61 ± 0.08	1.95 ± 0.06	2.20 ± 0.11	1.58
$N(0, 0.45)$	1.65 ± 0.08	1.85 ± 0.07	2.20 ± 0.09	2.58 ± 0.08	2.93 ± 0.12	3.30 ± 0.12	2.42
AVG	0.86	0.98	1.20	1.39	1.59	1.82	1.31

MNIST – TEST ERROR RATE [%] OF A 10-COMMITTEE

Feature noise	Label noise						AVG
	$p = 0$	$p = 0.05$	$p = 0.15$	$p = 0.25$	$p = 0.35$	$p = 0.45$	
-	0.50 ± 0.03	0.59 ± 0.03	0.71 ± 0.04	0.94 ± 0.04	0.97 ± 0.05	1.17 ± 0.04	0.81
$N(0, 0.05)$	0.49 ± 0.03	0.59 ± 0.03	0.75 ± 0.04	0.85 ± 0.03	1.05 ± 0.04	1.15 ± 0.04	0.81
$N(0, 0.15)$	0.54 ± 0.02	0.70 ± 0.04	0.88 ± 0.05	1.02 ± 0.04	1.16 ± 0.04	1.23 ± 0.05	0.92
$N(0, 0.25)$	0.72 ± 0.03	0.80 ± 0.04	1.07 ± 0.05	1.19 ± 0.04	1.30 ± 0.05	1.61 ± 0.07	1.12
$N(0, 0.35)$	0.96 ± 0.04	1.17 ± 0.06	1.35 ± 0.05	1.54 ± 0.05	1.83 ± 0.06	2.08 ± 0.09	1.49
$N(0, 0.45)$	1.51 ± 0.06	1.77 ± 0.06	2.12 ± 0.04	2.47 ± 0.08	2.79 ± 0.08	3.14 ± 0.09	2.30
AVG	0.79	0.94	1.15	1.34	1.52	1.73	1.24

MNIST – TEST ERROR RATE [%] OF A 15-COMMITTEE

Feature noise	Label noise						AVG
	$p = 0$	$p = 0.05$	$p = 0.15$	$p = 0.25$	$p = 0.35$	$p = 0.45$	
-	0.49 ± 0.03	0.58 ± 0.03	0.70 ± 0.03	0.92 ± 0.04	0.94 ± 0.04	1.15 ± 0.04	0.80
$N(0, 0.05)$	0.48 ± 0.03	0.57 ± 0.03	0.73 ± 0.04	0.83 ± 0.03	1.04 ± 0.03	1.13 ± 0.04	0.80
$N(0, 0.15)$	0.52 ± 0.03	0.68 ± 0.03	0.87 ± 0.03	1.01 ± 0.04	1.15 ± 0.03	1.21 ± 0.04	0.91
$N(0, 0.25)$	0.71 ± 0.03	0.79 ± 0.03	1.06 ± 0.05	1.16 ± 0.04	1.28 ± 0.05	1.58 ± 0.06	1.10
$N(0, 0.35)$	0.92 ± 0.04	1.16 ± 0.03	1.32 ± 0.04	1.53 ± 0.05	1.81 ± 0.04	2.05 ± 0.08	1.47
$N(0, 0.45)$	1.48 ± 0.05	1.73 ± 0.04	2.09 ± 0.04	2.44 ± 0.05	2.75 ± 0.06	3.10 ± 0.06	2.27
AVG	0.77	0.92	1.13	1.32	1.50	1.70	1.22

MNIST – TEST ERROR RATE [%] OF A 20-COMMITTEE

Feature noise	Label noise						AVG
	$p = 0$	$p = 0.05$	$p = 0.15$	$p = 0.25$	$p = 0.35$	$p = 0.45$	
-	0.48 ± 0.02	0.58 ± 0.03	0.69 ± 0.03	0.91 ± 0.03	0.94 ± 0.04	1.15 ± 0.04	0.79
$N(0, 0.05)$	0.48 ± 0.02	0.56 ± 0.02	0.72 ± 0.03	0.83 ± 0.03	1.03 ± 0.03	1.11 ± 0.03	0.79
$N(0, 0.15)$	0.51 ± 0.02	0.68 ± 0.03	0.87 ± 0.03	0.99 ± 0.04	1.15 ± 0.03	1.19 ± 0.04	0.90
$N(0, 0.25)$	0.71 ± 0.02	0.78 ± 0.02	1.05 ± 0.04	1.14 ± 0.03	1.28 ± 0.04	1.56 ± 0.05	1.09
$N(0, 0.35)$	0.91 ± 0.03	1.15 ± 0.03	1.32 ± 0.03	1.51 ± 0.04	1.79 ± 0.05	2.05 ± 0.06	1.46
$N(0, 0.45)$	1.45 ± 0.05	1.73 ± 0.04	2.09 ± 0.04	2.42 ± 0.04	2.74 ± 0.06	3.09 ± 0.05	2.25
AVG	0.76	0.91	1.12	1.30	1.49	1.69	1.21

Wyniki dla CIFAR-10

CIFAR-10 – TEST ERROR RATE [%] OF A SINGLE CNN

Feature noise	Label noise						AVG
	$p = 0$	$p = 0.05$	$p = 0.15$	$p = 0.25$	$p = 0.35$	$p = 0.45$	
-	21.46 ± 0.97	23.26 ± 0.69	25.70 ± 0.72	28.47 ± 0.60	31.52 ± 0.71	35.20 ± 0.80	27.57
$N(0, 0.05)$	24.09 ± 0.71	25.59 ± 0.59	28.24 ± 0.55	31.03 ± 0.50	33.80 ± 0.67	37.02 ± 0.72	29.92
$N(0, 0.15)$	32.14 ± 0.62	33.40 ± 0.66	35.73 ± 0.73	38.22 ± 0.71	40.70 ± 0.67	44.11 ± 0.85	37.37
$N(0, 0.25)$	38.99 ± 0.73	40.27 ± 0.64	42.46 ± 0.87	44.44 ± 0.57	47.20 ± 1.19	50.28 ± 0.74	43.88
$N(0, 0.35)$	45.55 ± 1.12	46.07 ± 0.70	47.97 ± 0.78	49.82 ± 0.58	52.75 ± 0.69	57.51 ± 8.71	49.90
$N(0, 0.45)$	49.60 ± 0.69	50.89 ± 0.53	52.43 ± 0.77	53.95 ± 0.60	56.13 ± 0.66	58.34 ± 0.73	53.52
AVG	35.25	36.52	38.72	40.95	43.67	47.05	40.36

CIFAR-10 – TEST ERROR RATE [%] OF A 5-COMMITTEE

Feature noise	Label noise						AVG
	$p = 0$	$p = 0.05$	$p = 0.15$	$p = 0.25$	$p = 0.35$	$p = 0.45$	
-	17.60 ± 0.39	19.18 ± 0.39	21.20 ± 0.39	23.67 ± 0.47	26.36 ± 0.50	29.50 ± 0.39	22.88
$N(0, 0.05)$	19.99 ± 0.39	21.47 ± 0.52	23.76 ± 0.40	25.99 ± 0.36	28.09 ± 0.36	31.70 ± 0.39	25.10
$N(0, 0.15)$	27.98 ± 0.32	28.99 ± 0.33	30.67 ± 0.40	32.89 ± 0.39	35.90 ± 0.48	38.75 ± 0.59	32.47
$N(0, 0.25)$	34.50 ± 0.34	36.02 ± 0.44	38.05 ± 0.53	40.12 ± 0.47	42.73 ± 0.84	45.39 ± 0.52	39.43
$N(0, 0.35)$	41.33 ± 0.68	42.48 ± 0.43	44.08 ± 0.55	45.77 ± 0.36	48.86 ± 0.41	51.40 ± 0.59	45.60
$N(0, 0.45)$	46.31 ± 0.47	47.09 ± 0.47	49.33 ± 0.53	50.71 ± 0.40	52.85 ± 0.59	54.95 ± 0.58	50.17
AVG	31.25	32.47	34.47	36.47	39.08	41.92	35.94

CIFAR-10 – TEST ERROR RATE [%] OF A 10-COMMITTEE

Feature noise	Label noise						AVG
	$p = 0$	$p = 0.05$	$p = 0.15$	$p = 0.25$	$p = 0.35$	$p = 0.45$	
-	17.07 ± 0.24	18.60 ± 0.38	20.64 ± 0.24	22.93 ± 0.38	25.61 ± 0.39	28.71 ± 0.25	22.23
$N(0, 0.05)$	19.46 ± 0.29	20.84 ± 0.30	23.19 ± 0.31	25.38 ± 0.25	27.22 ± 0.28	30.75 ± 0.31	24.42
$N(0, 0.15)$	27.39 ± 0.33	28.33 ± 0.26	29.91 ± 0.36	32.13 ± 0.29	35.05 ± 0.45	38.17 ± 0.42	31.78
$N(0, 0.25)$	33.83 ± 0.25	35.40 ± 0.33	37.32 ± 0.37	39.58 ± 0.32	41.88 ± 0.40	44.71 ± 0.37	38.75
$N(0, 0.35)$	40.62 ± 0.53	41.83 ± 0.37	43.48 ± 0.33	45.18 ± 0.34	48.21 ± 0.35	50.79 ± 0.37	44.97
$N(0, 0.45)$	45.81 ± 0.25	46.62 ± 0.30	48.97 ± 0.34	50.27 ± 0.34	52.45 ± 0.38	54.41 ± 0.44	49.72
AVG	30.65	31.92	33.87	35.85	38.37	41.22	35.31

CIFAR-10 – TEST ERROR RATE [%] OF A 15-COMMITTEE

Feature noise	Label noise						AVG
	$p = 0$	$p = 0.05$	$p = 0.15$	$p = 0.25$	$p = 0.35$	$p = 0.45$	
-	16.93 ± 0.16	18.42 ± 0.32	20.37 ± 0.23	22.72 ± 0.31	25.32 ± 0.32	28.50 ± 0.27	22.02
$N(0, 0.05)$	19.31 ± 0.21	20.66 ± 0.26	22.95 ± 0.25	25.12 ± 0.21	26.97 ± 0.18	30.53 ± 0.30	24.22
$N(0, 0.15)$	27.20 ± 0.33	28.09 ± 0.18	29.67 ± 0.24	31.79 ± 0.26	34.87 ± 0.26	37.86 ± 0.27	31.52
$N(0, 0.25)$	33.62 ± 0.22	35.12 ± 0.26	37.05 ± 0.32	39.35 ± 0.31	41.64 ± 0.29	44.51 ± 0.23	38.52
$N(0, 0.35)$	40.40 ± 0.36	41.56 ± 0.27	43.19 ± 0.30	44.92 ± 0.26	48.03 ± 0.27	50.60 ± 0.29	44.75
$N(0, 0.45)$	45.58 ± 0.24	46.43 ± 0.20	48.78 ± 0.28	50.10 ± 0.25	52.26 ± 0.26	54.18 ± 0.33	49.50
AVG	30.48	31.67	33.60	35.63	38.13	41.00	35.09

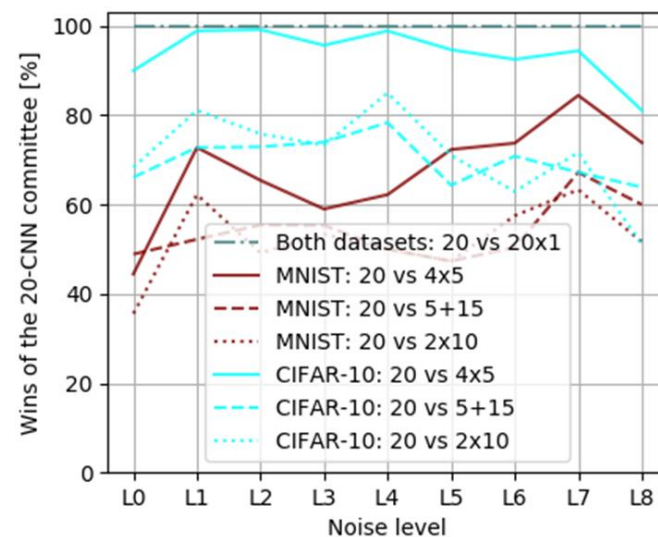
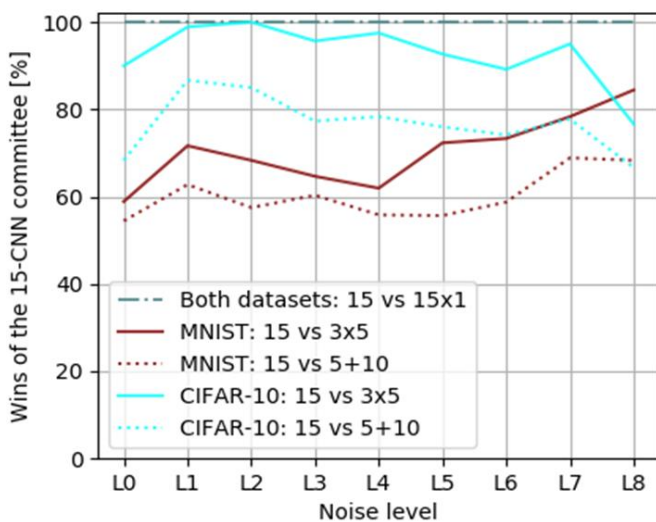
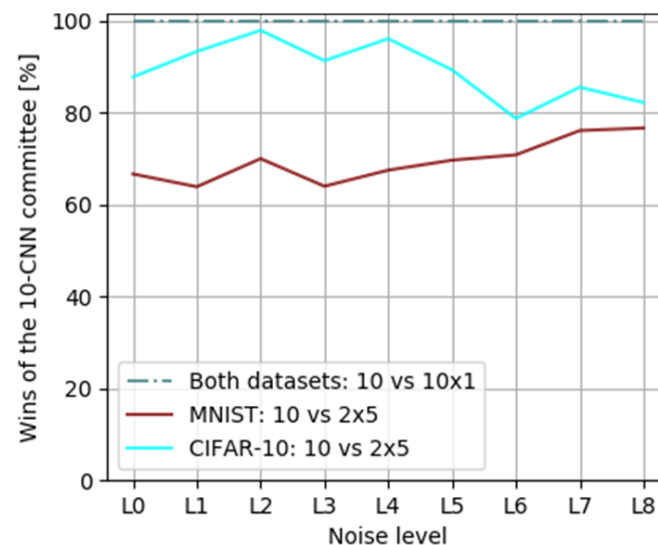
CIFAR-10 – TEST ERROR RATE [%] OF A 20-COMMITTEE

Feature noise	Label noise						AVG
	$p = 0$	$p = 0.05$	$p = 0.15$	$p = 0.25$	$p = 0.35$	$p = 0.45$	
-	16.86 ± 0.16	18.30 ± 0.27	20.32 ± 0.19	22.60 ± 0.24	25.15 ± 0.24	28.29 ± 0.20	21.88
$N(0, 0.05)$	19.15 ± 0.20	20.51 ± 0.19	22.80 ± 0.17	25.07 ± 0.19	26.87 ± 0.18	30.34 ± 0.27	24.08
$N(0, 0.15)$	27.04 ± 0.24	27.98 ± 0.19	29.49 ± 0.19	31.57 ± 0.21	34.71 ± 0.25	37.74 ± 0.21	31.37
$N(0, 0.25)$	33.53 ± 0.18	35.01 ± 0.22	36.88 ± 0.26	39.24 ± 0.25	41.52 ± 0.22	44.34 ± 0.15	38.38
$N(0, 0.35)$	40.32 ± 0.34	41.46 ± 0.22	43.07 ± 0.19	44.82 ± 0.21	48.00 ± 0.25	50.54 ± 0.23	44.67
$N(0, 0.45)$	45.41 ± 0.21	46.43 ± 0.17	48.71 ± 0.23	49.94 ± 0.24	52.15 ± 0.23	54.15 ± 0.27	49.43
AVG	30.35	31.58	33.50	35.50	38.03	40.85	34.97

Różne konfiguracje CNNów

FEATURE-LABEL NOISE LEVELS

Feature noise	Label noise					
	$p = 0$	$p = 0.05$	$p = 0.15$	$p = 0.25$	$p = 0.35$	$p = 0.45$
-	L0	L0	L1	L2	L3	L4
$N(0, 0.05)$	L0	L1	L2	L3	L4	L5
$N(0, 0.15)$	L1	L2	L3	L4	L5	L6
$N(0, 0.25)$	L2	L3	L4	L5	L6	L7
$N(0, 0.35)$	L3	L4	L5	L6	L7	L8
$N(0, 0.45)$	L4	L5	L6	L7	L8	L8





Podsumowanie wyników

- Średni błąd oraz jego odchylenie standardowe maleją wraz ze wzrostem rozmiaru komitetu
- Prędkość poprawy stopniowo maleje
- Komitety zawsze pokonują zbiór pojedynczych modeli, niezależnie od stopnia zaszumienia i wielkości zespołu
- Duże komitety pokonują kombinacje mniejszych dla większości stopni zaszumienia
- Duże komitety radzą sobie relatywnie lepiej dla zbioru CIFAR-10 niż MNIST
- Nie stwierdzono jasnej relacji pomiędzy poziomem szumu a optymalną konfiguracją modeli

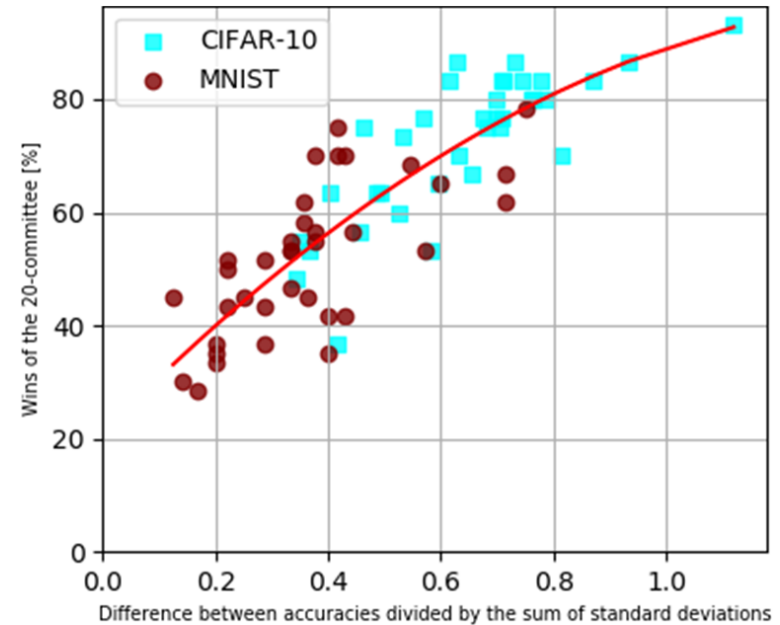
Różnica średnich a suma wariancji

TEST ERROR RATE [%] OF A 10-COMMITTEE

Feature noise	Label noise			
	$p = 0$	$p = 0.05$	$p = 0.15$	$p = 0.25$
-	17.07 ± 0.24	18.60 ± 0.38	20.64 ± 0.24	22.93 ± 0.38
$N(0, 0.05)$	19.46 ± 0.29	20.84 ± 0.30	23.19 ± 0.31	25.38 ± 0.25
$N(0, 0.15)$	27.39 ± 0.33	28.33 ± 0.26	29.91 ± 0.36	32.13 ± 0.29
$N(0, 0.25)$	33.83 ± 0.25	35.40 ± 0.33	37.32 ± 0.37	39.58 ± 0.32

TEST ERROR RATE [%] OF A 20-COMMITTEE

Feature noise	Label noise			
	$p = 0$	$p = 0.05$	$p = 0.15$	$p = 0.25$
-	16.86 ± 0.16	18.30 ± 0.27	20.32 ± 0.19	22.60 ± 0.24
$N(0, 0.05)$	19.15 ± 0.20	20.51 ± 0.19	22.80 ± 0.17	25.07 ± 0.19
$N(0, 0.15)$	27.04 ± 0.24	27.98 ± 0.19	29.49 ± 0.19	31.57 ± 0.21
$N(0, 0.25)$	33.53 ± 0.18	35.01 ± 0.22	36.88 ± 0.26	39.24 ± 0.25



20-committee performance against two 10-committees.

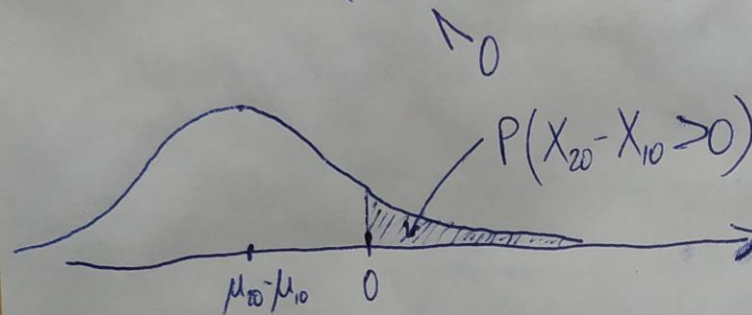
Różnica średnich a suma wariancji

$X_{10} \sim N(\mu_{10}, \sigma_{10}^2)$ - miernik losowe błędów dla 10-komitetu, np. $N(0.51, 0.02^2)$
 $X_{20} \sim N(\mu_{20}, \sigma_{20}^2)$ - ————— " ————— 20-komitetu, np. $N(0.48, 0.03^2)$

Zastanawiamy się kiedy $P(X_{20} > X_{10}) = P(X_{20} - X_{10} > 0)$ jest największe.

Jest to prawd. że jeden 10-komitet pólno jeden 20-komitet.
 (tylki błąd 20-komitetu > błąd 10-komitetu)

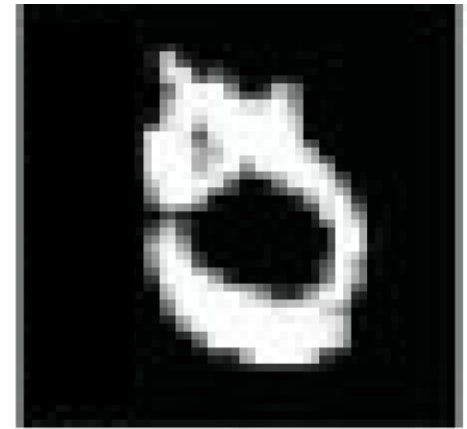
$$X_{20} - X_{10} \sim N(\mu_{20} - \mu_{10}, \sigma_{20}^2 + \sigma_{10}^2)$$



Zatem, aby $P(X_{20} - X_{10} > 0)$ było jak największe, $\sigma_{20}^2 + \sigma_{10}^2$ też musi być jak największe

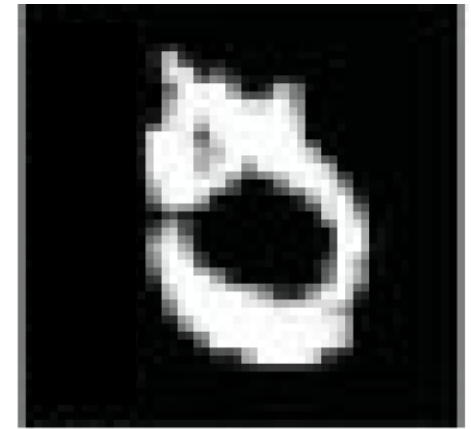
Analiza błędów

- Błędnie/wątpliwie poetykietowane egzemplarze zbioru testowego



Analiza błędów

- Błędnie/wątpliwie poetykietowane egzemplarze zbioru testowego



8

Analiza błędów

- Błędnie/wątpliwie poetykietowane egzemplarze zbioru testowego



8



9



Analiza błędów

- Błędnie/wątpliwie poetykietowane egzemplarze zbioru testowego



8



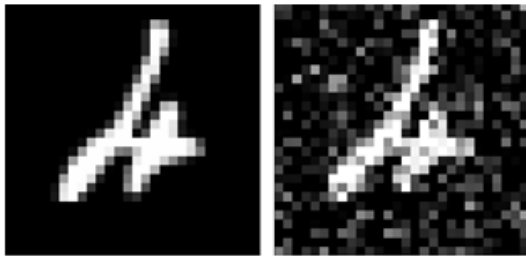
9



5

Analiza błędów

- Jeden błędnie klasyfikujący model, który mocno wpływa na całościową predykcję

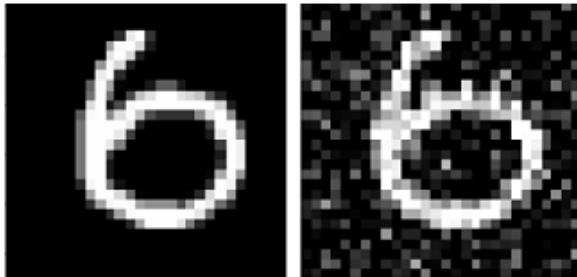


- obraz z szumem gaussowskim $N(0, 0.25)$
 - 25% błędnych etykiet w procesie uczenia

Class	Probability vectors										SUM (V1:V5)	SUM (V1:V10)
	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10		
Digit 0	0.042	0.032	0.043	0.030	0.041	0.032	0.025	0.029	0.015	0.034	0.188	0.323
Digit 1	0.094	0.051	0.115	0.094	0.093	0.033	0.102	0.099	0.040	0.101	0.447	0.822
Digit 2	0.116	0.140	0.175	0.142	0.088	0.219	0.103	0.061	0.119	0.062	0.661	1.225
Digit 3	0.052	0.045	0.064	0.042	0.032	0.022	0.052	0.046	0.015	0.041	0.235	0.411
Digit 4	0.174	0.324	0.221	0.122	0.338	0.425	0.395	0.163	0.454	0.251	1.179	2.867
Digit 5	0.046	0.048	0.040	0.028	0.048	0.031	0.014	0.033	0.016	0.050	0.210	0.354
Digit 6	0.281	0.220	0.174	0.418	0.211	0.077	0.191	0.368	0.211	0.310	1.304	2.461
Digit 7	0.052	0.042	0.062	0.057	0.020	0.008	0.044	0.042	0.079	0.046	0.233	0.452
Digit 8	0.110	0.073	0.071	0.047	0.107	0.140	0.045	0.121	0.040	0.073	0.408	0.827
Digit 9	0.033	0.025	0.036	0.020	0.022	0.013	0.028	0.038	0.012	0.033	0.136	0.260

Analiza błędów

- Ciężko jednoznacznie stwierdzić przyczynę błędu



- obraz z szumem gaussowskim $N(0, 0.25)$
- 5% błędnych etykiet w procesie uczenia

Class	Probability vectors										SUM (V1:V5)	SUM (V1:V10)
	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10		
Digit 0	0.346	0.591	0.591	0.055	0.792	0.838	0.650	0.345	0.338	0.585	2.375	5.131
Digit 1	0.017	0.016	0.023	0.007	0.020	0.009	0.011	0.024	0.088	0.050	0.083	0.265
Digit 2	0.001	0.001	0.004	0.002	0.001	0.001	0.002	0.006	0.013	0.007	0.009	0.038
Digit 3	0.001	0.002	0.001	0.001	0.000	0.003	0.003	0.008	0.004	0.006	0.005	0.029
Digit 4	0.001	0.003	0.002	0.007	0.002	0.001	0.002	0.006	0.006	0.006	0.015	0.036
Digit 5	0.006	0.004	0.082	0.090	0.146	0.007	0.003	0.022	0.029	0.094	0.328	0.483
Digit 6	0.619	0.355	0.274	0.806	0.024	0.133	0.320	0.553	0.469	0.219	2.078	3.772
Digit 7	0.004	0.009	0.016	0.005	0.006	0.002	0.002	0.012	0.013	0.012	0.040	0.081
Digit 8	0.004	0.012	0.005	0.015	0.003	0.003	0.004	0.024	0.031	0.011	0.039	0.112
Digit 9	0.001	0.007	0.003	0.012	0.006	0.003	0.004	0.002	0.007	0.009	0.029	0.054

- Przy tym samym stopniu zaszumienia obrazu, ale bez zaszumienia etykiet, komitet silnie wskazuje na liczbę sześć, np. dla 10-komitetu skumulowane prawdopodobieństwo zera wynosi 1.818, a szóstki – 8.166
- 5% zaszumionych etykiet drastycznie wpływa na wynik



Podsumowanie i dalsze prace

- Zespoły modeli są dobre!
- Bardziej złożone architektury pojedynczych modeli
- Początkowe odszumienie
- Augmentacja danych
- Zwiększenie niezależności modeli wchodzących w skład komitetów
- Zastosowanie komitetów do innych problemów klasyfikacji



Źródła

1. R. J. Hickey, “Noise modelling and evaluating learning from examples”, *Artif. Intell.*, vol. 82, no. 1-2, pp. 157–179, 1996.
2. J. R. Quinlan, “Induction of decision trees”, *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986.
3. B. Frenay and M. Verleysen, "Classification in the Presence of Label Noise: A Survey" in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 5, pp. 845-869, May 2014.
4. X. Zhu and X. Wu, “Class noise vs. attribute noise: A quantitative study”, *Artif. Intell. Rev.*, vol. 22, pp. 177–210, 2004.
5. <https://adventuresinmachinelearning.com/keras-tutorial-cnn-11-lines/>