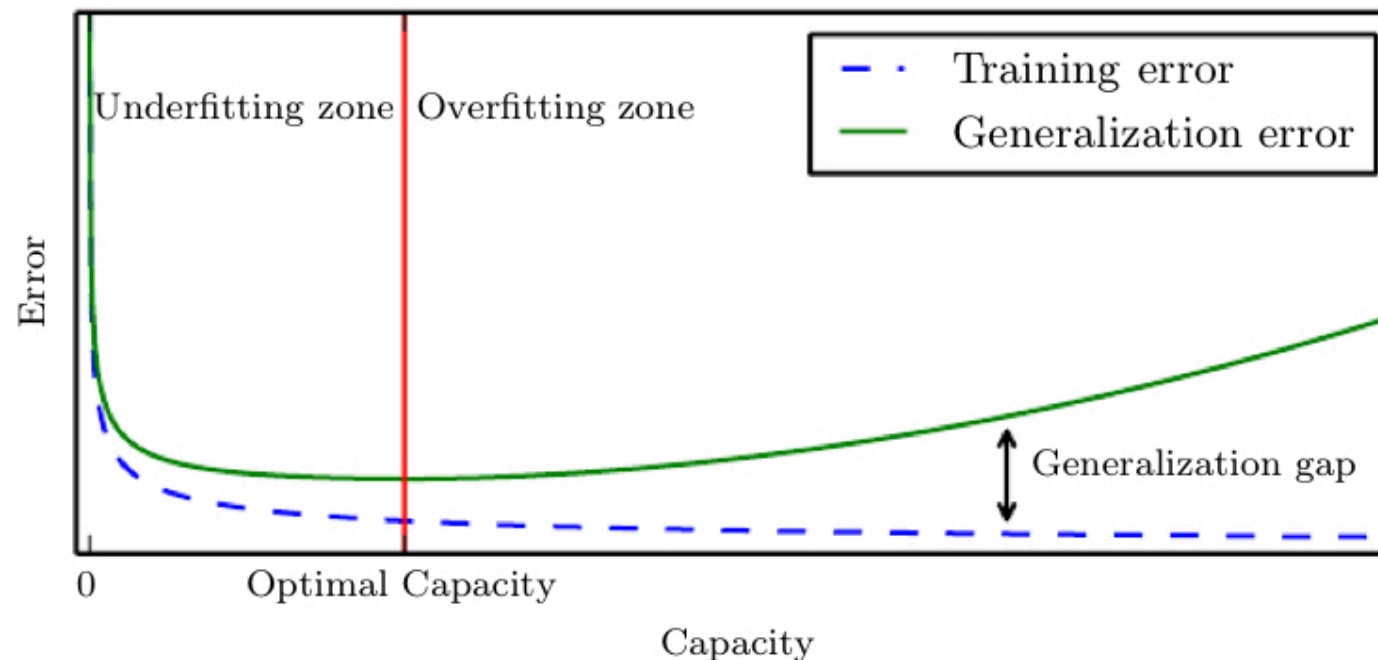


Wykład 7

Przygotowanie danych i
regularyzacja

Rodzaje błędów



- Błąd treningu
- Błąd generalizacji (testu)
- Pojemność - zdolność modelu do odwzorowania szerokiego wachlarza funkcji

Regularyzacja

Definicja

Strategie używane w nauczaniu maszynowym mające na celu redukcję błędu na zbiorze testowym (często kosztem błędu treningu).

Norma parametrów modelu

$$\tilde{J}(\boldsymbol{\theta}; \boldsymbol{X}, \boldsymbol{y}) = J(\boldsymbol{\theta}; \boldsymbol{X}, \boldsymbol{y}) + \alpha \Omega(\boldsymbol{\theta})$$

$$\tilde{J}(\boldsymbol{w}; \boldsymbol{X}, \boldsymbol{y}) = \frac{\alpha}{2} \boldsymbol{w}^\top \boldsymbol{w} + J(\boldsymbol{w}; \boldsymbol{X}, \boldsymbol{y}),$$

finding parameter gradient

Dla normalizacji w L2

$$\nabla_{\boldsymbol{w}} \tilde{J}(\boldsymbol{w}; \boldsymbol{X}, \boldsymbol{y}) = \alpha \boldsymbol{w} + \nabla_{\boldsymbol{w}} J(\boldsymbol{w}; \boldsymbol{X}, \boldsymbol{y}).$$

gradient step to update the weights, we perform

$$\Omega(\boldsymbol{\theta}) = \frac{1}{2} \|\boldsymbol{w}\|_2^2$$

$$\boldsymbol{w} \leftarrow \boldsymbol{w} - \epsilon (\alpha \boldsymbol{w} + \nabla_{\boldsymbol{w}} J(\boldsymbol{w}; \boldsymbol{X}, \boldsymbol{y})).$$

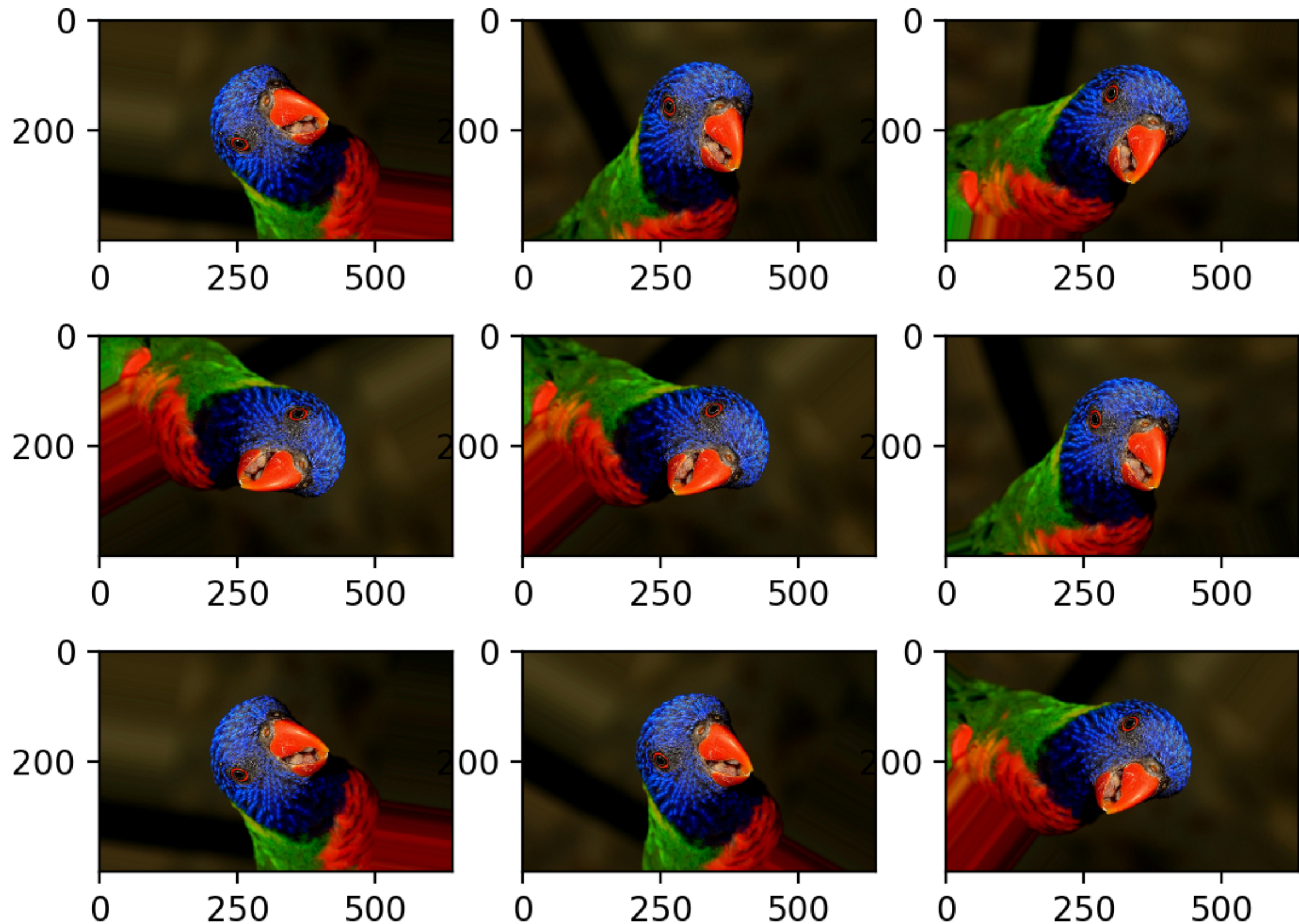
in other words, the update is:

$$\boldsymbol{w} \leftarrow (1 - \epsilon \alpha) \boldsymbol{w} - \epsilon \nabla_{\boldsymbol{w}} J(\boldsymbol{w}; \boldsymbol{X}, \boldsymbol{y}).$$

Modyfikacje zbioru danych

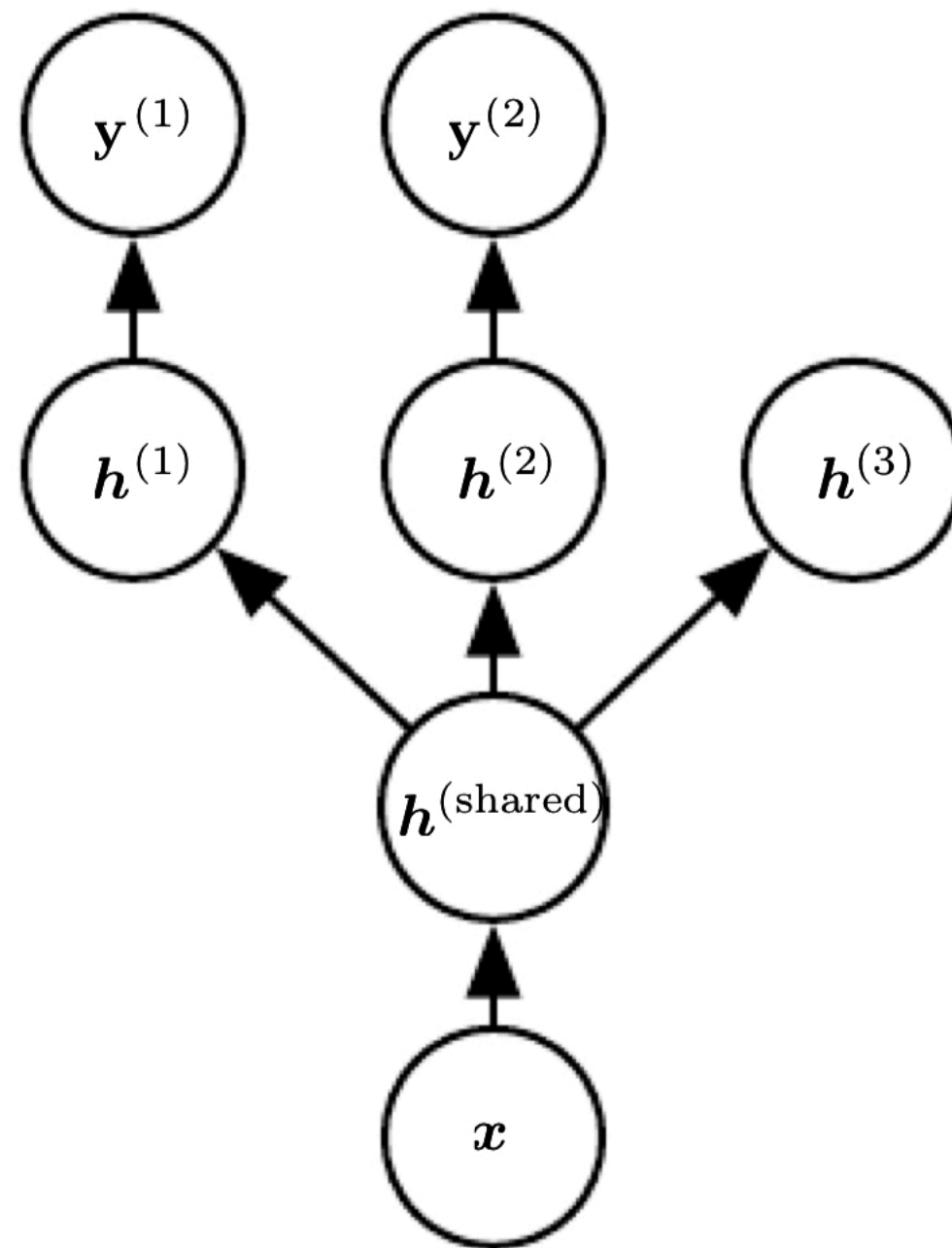
Więcej danych!

- Transformacja danych wejściowych (klasyfikacja)
- “sztuczne” dane (projekt: tabelki)
- Obroty, przesunięcia, dodawanie szumów
- Noise on labels



<https://machinelearningmastery.com/how-to-configure-image-data-augmentation-when-training-deep-learning-neural-networks/>

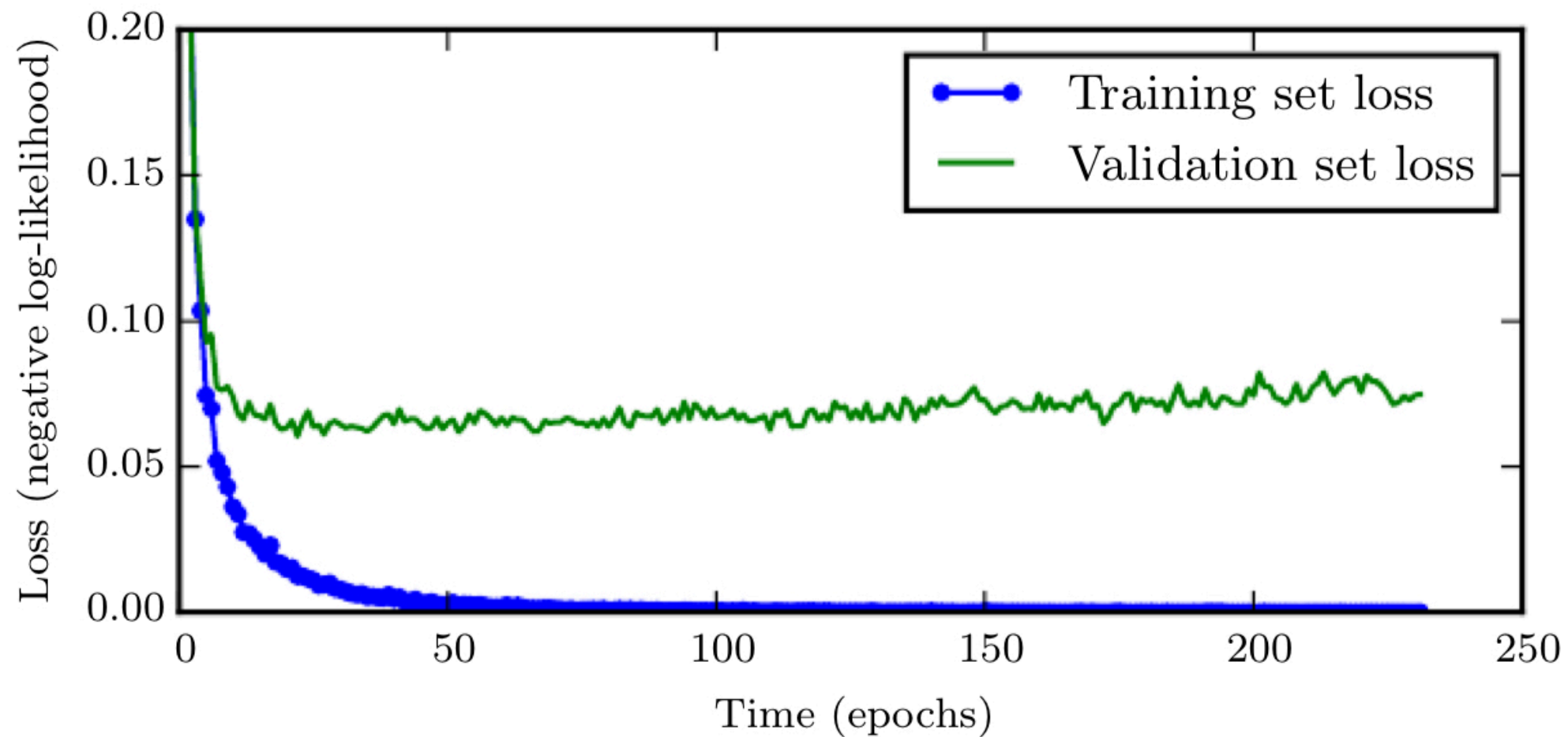
Multi-task learning



Multi-task vs transfer learning:

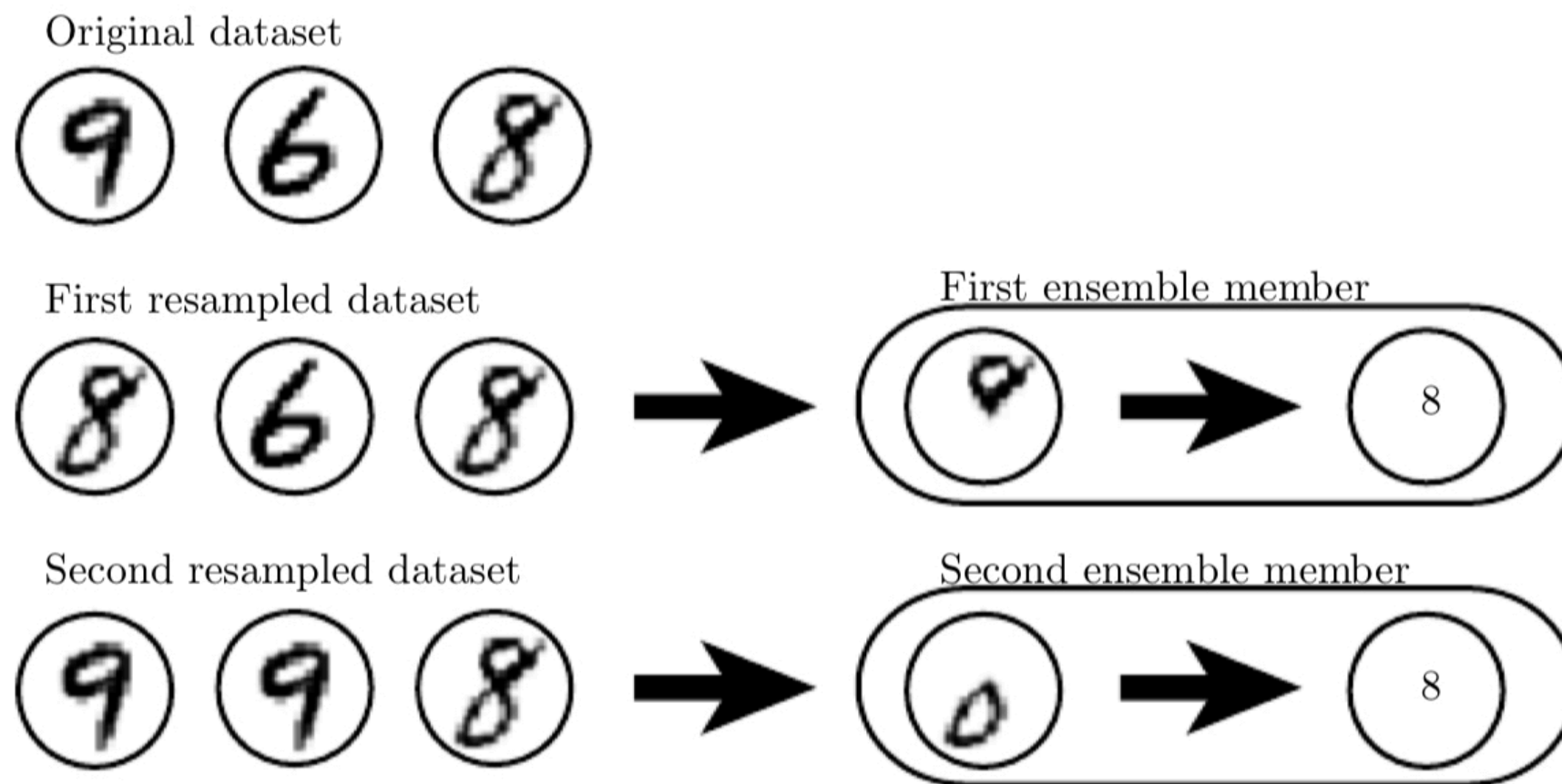
<https://www.linkedin.com/pulse/transfer-learning-vs-multitask-ibrahim-sobh/>

Early stopping

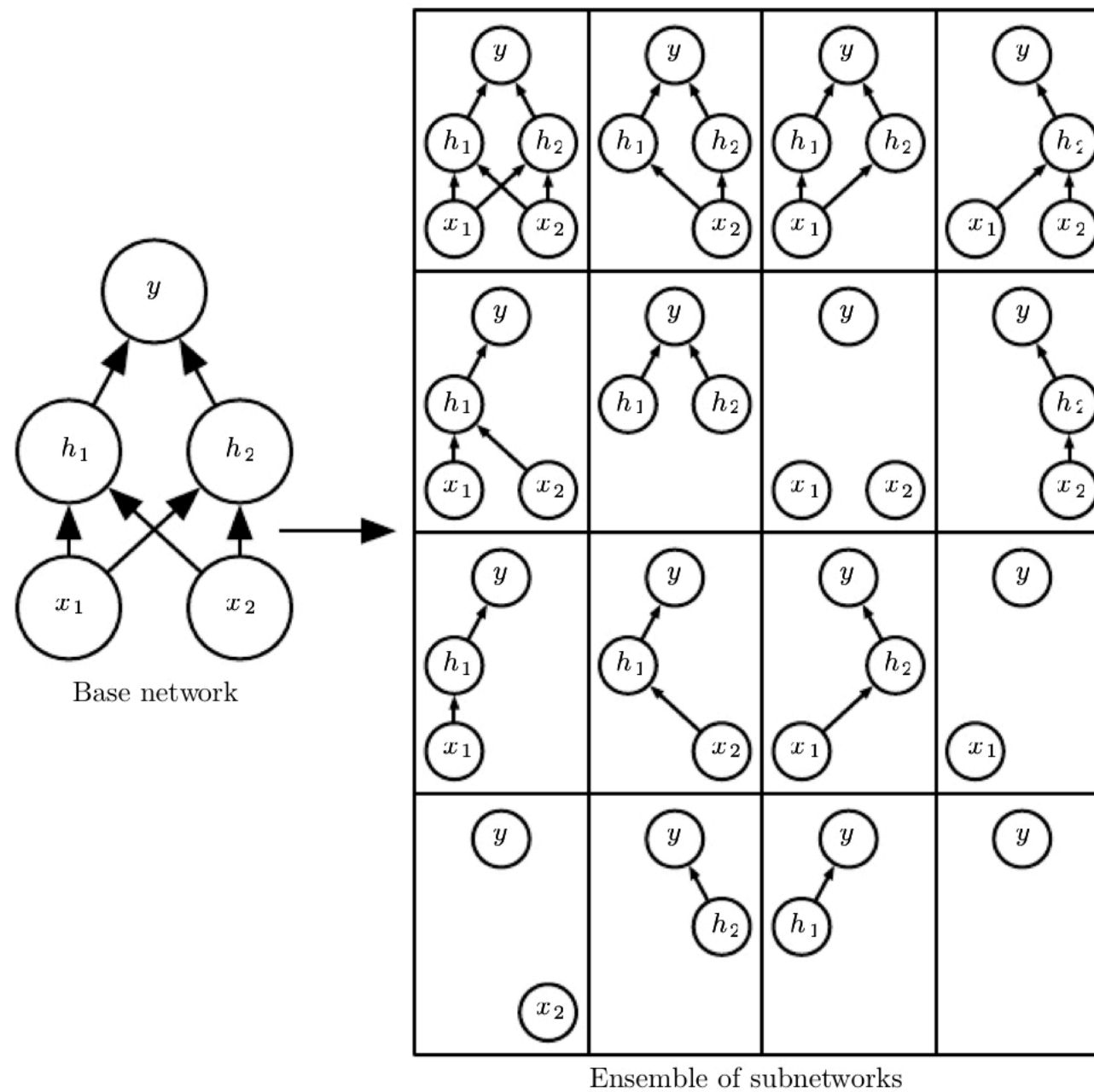


Składanie modeli

Bagging = Bootstrap aggregating



Wyłączanie neuronów (dropout)



Adversarial training



x

$y = \text{"panda"}$
w/ 57.7%
confidence

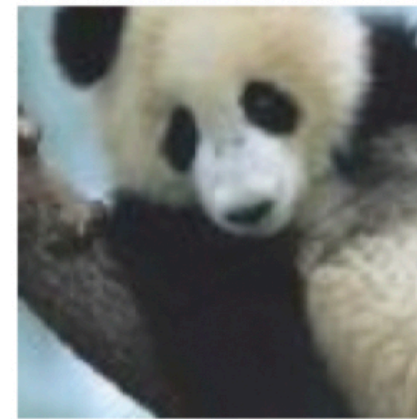
$+ .007 \times$



$\text{sign}(\nabla_x J(\theta, x, y))$

"nematode"
w/ 8.2%
confidence

$=$



$x +$
 $\epsilon \text{sign}(\nabla_x J(\theta, x, y))$
"gibbon"
w/ 99.3 %
confidence

