



deeplearning.ai

Face recognition

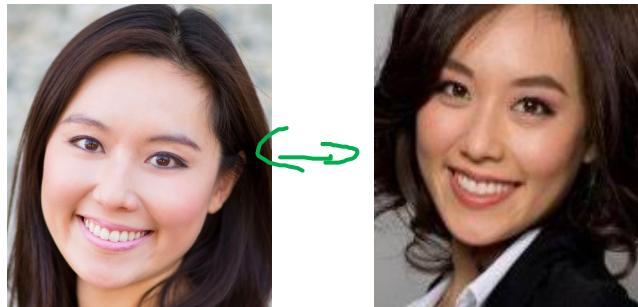
Triplet loss

NN가 얼굴 사진에 대해 good encoding을 만들도록 NN의 파라미터를 학습하는 한가지 방법은 triplet loss function에 gradient descent를 적용하는 것이다.

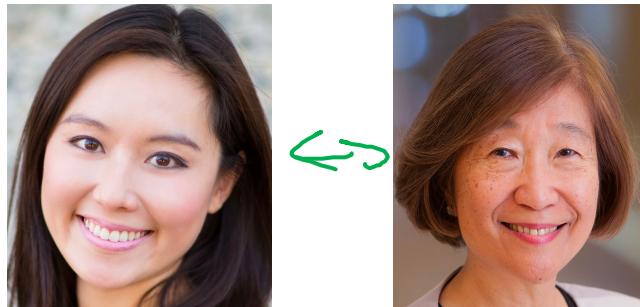
triplet loss function을 이용해 NN의 파라미터를 학습하기 위해서는 여러 이미지를 동시에 봐야한다.

Learning Objective

예를 들어 this pair of images가 주어졌을 때 encoding이 유사하기를 원한다.
왜냐하면 동일인이니까.



반면 얘들의 encoding은 다르기를 원한다.



triplet loss라는 이름이 붙은 이유는 항상 앵커 이미지 하나에 대해서 positive image는 거리가 가깝게, negative image는 거리를 멀게 만들기를 원하기 때문이다.
(항상 동시에 이미지 3장을 처리한다)

$$\begin{array}{c} \text{Anchor} \\ \hline A \end{array} \quad \begin{array}{c} \text{Positive} \\ \hline P \end{array}$$

$d(A, P) = 0.5$

고로 이 margin(gap)은 $d(A, P)$ 는 push down 해주고 $d(A, N)$ 은 push up 해줘서 그 둘 간을 훨씬 벌려주는 효과가 되는 것

Want : $\|f(A) - f(P)\|^2 + \lambda \leq 0$

what you want is...
encoding이 다음과 같은 속성을 가지도록
NN의 파라미터가 학습되기를 원한다.

$$\begin{array}{c} \text{Anchor} \\ \hline A \end{array} \quad \begin{array}{c} \text{Negative} \\ \hline N \end{array}$$

$d(A, N) = 0.7$

$$\|f(A) - f(N)\|^2 \leq d(A, N)$$

위 식을 다시 쓰면

$$\|f(A) - f(P)\|^2 - \|f(A) - f(N)\|^2 + \lambda \leq 0$$

Margin

예를 들어 이 경우, 0.51은 positive pair의 거리 0.5보다 크긴 하지만 이걸로 충분치는 않다고 $d(A, N)$ 이 $d(A, P)$ 보다 좀 더 많이 크기를 원한다고 말해주는 효과가 있다. 적어도 0.7은 돼야지...

모든 이미지에 대해서 output을 0으로 만들도록 f 가 학습이 되면 왼쪽 식을 항상 만족하게 되는데 이건 우리가 원하는게 아니다.

$$f(\text{img}) = \vec{0}$$

objective를 약간 수정하면 된다. 0보다 작거나 같기만 하면 되는게 아니라 0보다 약간 작은 수 보다 작아야 한다고 주장

Andrew Ng

Loss function

Given 3 images

A, P, N :

$$\underline{L(A, P, N)} = \max \left(\left[\|f(A) - f(P)\|^2 - \|f(A) - f(N)\|^2 + \lambda \right], 0 \right)$$

$$J = \sum_{i=1}^m L(A^{(i)}, P^{(i)}, N^{(i)})$$

A, P
 T

Training set: $\underbrace{10k}_{\infty}$ pictures of $\frac{1k}{\infty}$ persons

Choosing the triplets A,P,N



During training, if A,P,N are chosen randomly,
 $d(A, P) + \alpha \leq d(A, N)$ is easily satisfied.

$$\underbrace{\|f(A) - f(P)\|^2}_{\text{Distance between } A \text{ and } P} + \alpha \leq \underbrace{\|f(A) - f(N)\|^2}_{\text{Distance between } A \text{ and } N}$$

Choose triplets that're “hard” to train on.

$$\begin{aligned} \cancel{d(A, P)} + \alpha &\leq d(A, N) \\ \frac{d(A, P)}{\downarrow} &\approx \frac{d(A, N)}{\uparrow} \end{aligned}$$

FaceNet
DeepFace



Training set using triplet loss

Anchor



Positive



Negative



:

:

:



J

$$d(x^{(i)}, x^{(j)})$$