deeplearning.ai

Face recognition

Triplet loss

This is the loss func: on which we apply SGD for facial recognition

Learning Objective

Triplet comes from the fact that you have 3 terms (Anchor, positive, Negative)

same person

Anchor (A)

Positive (P)

diff person

Anchor (A)

Negative (N)

- To Apply triplet loss, you need to compare pairs of Images
- We compare 2 pairs (A,P), (A,N)
  - where we want the encodings of A,P to be similar so that $\|f(A)-f(P)\|^2$ is small

  & we want encodings of A,N to be diff so that $\|f(A)-f(N)\|^2$ are large

$\Rightarrow \|f(A)-f(P)\|^2 \le \|f(A)-f(N)\|^2$ (or $d(A,P) \le d(A,N)$)

$\Rightarrow d(A,P) - d(A,N) \le 0$ → one way to do this is make $f(A)=f(P)=f(N)=$ constant (but we dont want the encoding to always output a constant)

∴ to Make sure that we have Some gap b/w $d(A,P)$ & $d(A,N)$

we do $d(A,P)-d(A,N) \le 0-\alpha$ or $\left[d(A,P)-d(A,N)+\alpha \le 0\right]$

α = Margin

[Schroff et al.,2015, FaceNet: A unified embedding for face recognition and clustering]

Loss function

Given 3 Images (A,P,N) → P = Image of same person
N = Image of diff person

$$L(A,P,N) = \max\left(\|f(A) - f(P)\|^2 - \|f(A) - f(N)\|^2 + \alpha, 0 \right)$$

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

Then generate M rows of 3 Images each (A,P,N) & train NN on each of these triplets.

Training set: 10k pictures of 1k persons

Need to have duplicate pictures so you get (A,P)

[Schroff et al.,2015, FaceNet: A unified embedding for face recognition and clustering]

Choosing the triplets A,P,N

(A, P = same person
A, N = diff persons)

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

If you choose at Random, chances that A,N are very diff is very probable

Choose triplets that're "hard" to train on.

choose triplets where

$$d(A,P) \approx d(A,N)$$

- If you choose Randomly, then a lot of triplets will be automatically satisfied w/o much change in params ⇒ gradient descent wouldnt do much learning! , "the real learning happens when things are hard to dicepher"

[Schroff et al.,2015, FaceNet: A unified embedding for face recognition and clustering]

Training set using triplet loss

Anchor	Positive	Negative
⋮	⋮	⋮