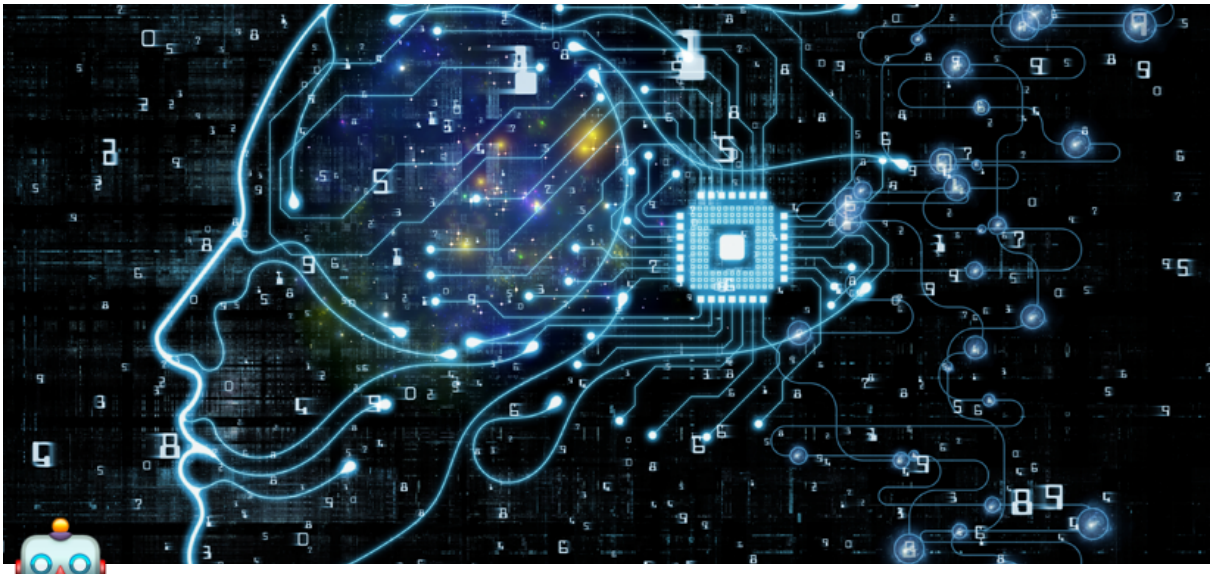


eLDA 보고서

code 기여도

1위 이인규

2위 류제경

레포트 작성 기여도

1위 류제경

2위 이인규

서론

기존의 선형 판별 분석(Linear Discriminant Analysis, LDA)은 특이성 문제, 차원의 저주, 비선형성 문제 등 여러 가지 한계를 내포하고 있습니다.

특히, **동적 데이터**에 대한 처리에서 두드러지는 약점을 보이며, 이는 음성 데이터 분야에서 더욱 명확하게 나타납니다.

음성 데이터의 복잡성과 변동성으로 인해 기존 LDA의 적용 사례는 매우 제한적이며, 실질적인 효과를 발휘하기 어려운 상황입니다.

이러한 문제를 해결하기 위해 본 보고서에서는 **지수 분포**를 기반으로 한 새로운 LDA 변형 기법인 지수형 선형 판별 분석(Exponential Linear Discriminant Analysis, eLDA)을 제안합니다.

eLDA는 지수 분포를 사용하여 LDA를 변형하고, 가중치 w를 도입함으로써 이상치에 대한 저항성을 강화하였습니다.

이 새로운 지수 분포 기반 접근법을 통해 음성 데이터의 특성을 보다 효과적으로 반영할 수 있게 됩니다.

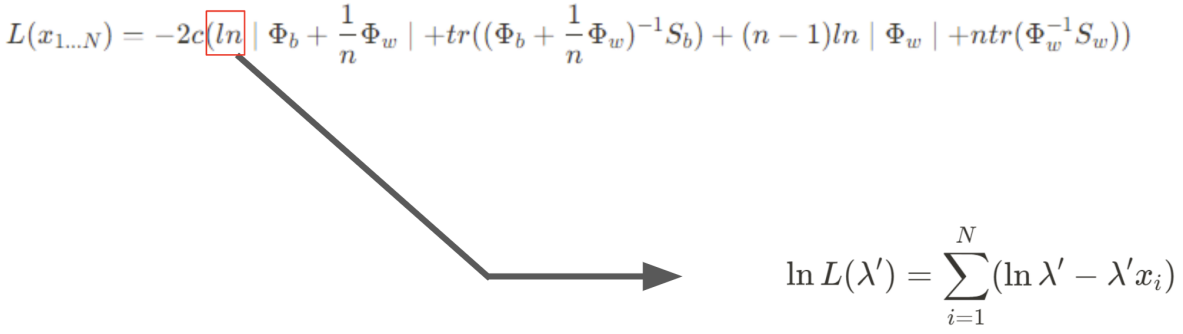
본 연구에서는 기존 LDA와 eLDA를 동일한 데이터 셋에 적용하여 정확도를 비교 분석함으로써 eLDA의 우수성을 입증하고자 합니다.

본 보고서는 음성 데이터 분석에 있어서 eLDA의 적용 가능성과 그 성능을 중심으로, 기존 방법론의 한계를 극복하는 새로운 접근법을 제시합니다.

이를 통해 eLDA가 음성 데이터 분석의 실효성을 높이고, 보다 정교한 분석을 가능하게 하는 잠재력을 보여줍니다.

관련연구

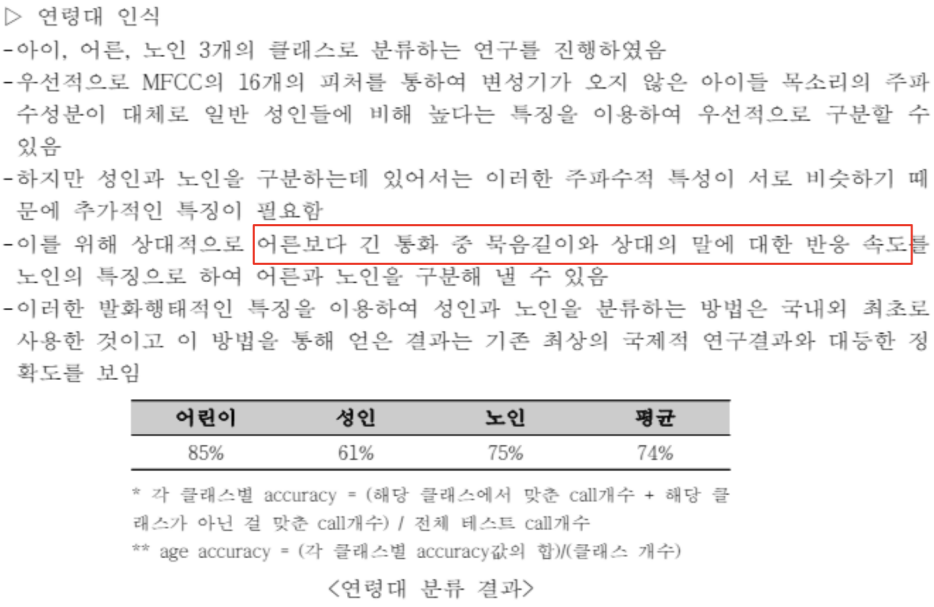
Probabilistic Linear Discriminant Analysis, European Conference on Computer Vision (ECCV), 2006.



이 논문에서 확률적 LDA를 다루기 위해 maximum likelihood를 목적함수로 하고 이를 최대화하는 방향으로 학습을 진행하였는데, 이를 착안해서 우리의 모델도 지수분포의 log likelihood를 구하고, 이를 목적함수로 설정하였다.

무음구간을 이용한 분류에 대한 타당성

한국과학기술원, 음성·음향 분석 기반 상황 판단 솔루션 기술 개발(2018)



▷ 연령대 인식
-아이, 어른, 노인 3개의 클래스로 분류하는 연구를 진행하였음
-우선적으로 MFCC의 16개의 피처를 통하여 변성기가 오지 않은 아이들 목소리의 주파수성분이 대체로 일반 성인들에 비해 높다는 특징을 이용하여 우선적으로 구분할 수 있음
-하지만 성인과 노인을 구분하는데 있어서는 이러한 주파수적 특성이 서로 비슷하기 때문에 추가적인 특징이 필요함
-이를 위해 상대적으로 어른보다 긴 통화 중 묵음길이와 상대의 말에 대한 반응 속도를 노인의 특징으로 하여 어른과 노인을 구분해 낼 수 있음
-이러한 발화행태적인 특징을 이용하여 성인과 노인을 분류하는 방법은 국내외 최초로 사용한 것이고 이 방법을 통해 얻은 결과는 기존 최상의 국제적 연구결과와 대등한 정확도를 보임

어린이	성인	노인	평균
85%	61%	75%	74%

* 각 클래스별 accuracy = (해당 클래스에서 맞춘 call개수 + 해당 클래스가 아닌 걸 맞춘 call개수) / 전체 테스트 call개수
** age accuracy = (각 클래스별 accuracy값의 합)/(클래스 개수)

<연령대 분류 결과>

음성·음향 분석 기반 상황 판단 솔루션 기술 개발(2018) 보고서 중 연령대 인식에 대한 설명

위 사진은 한국과학기술원에서 음성 음향 분석 기반 상환 판단 솔루션 기술 개발이란 보고서에서 나온 내용의 일부를 발췌한것이다. 이때, 어른과 노인의 묵음 길이에 대해 분석하여 구분해낼수있다고 함.

데이터 설명

유튜브 영상에서 소리만 따로 추출해 데이터를 만들었습니다.

1. 훈련 데이터

노인 데이터

김진홍 목사, 82세, 남성

https://youtu.be/5X4czzXGriw?si=CEStR1cqlfKhg8kD

https://youtu.be/pX_ofGhwJk4?si=g1LM-lH6Jx_U9noY

https://youtu.be/bvDsYXTrXWc?si=qiF9TcV9HBbTLNTu

위 영상들에서 5분단위로 끊어서 랜덤으로 묵음구간 500개를 이용해 훈련 데이터를 만들었습니다.

folder : train/OB/1~9 의 9개의 음원 파일

청년 데이터

나동빈, 30세, 남성

https://youtu.be/m-9pAwq1o3w?si=kZcb5rde1D4knrvf

https://youtu.be/AVvlDmhHgC4?si=zLfcBJDCPe4D69H1

역시, 노인데이터와같이, 5분단위로 끊어, 500개의 묵음구간을 선택해 훈련데이터로 사용했습니다.

folder : train/YB/1~6 의 6개의 음원 파일

2. 테스트 데이터

노인 데이터

김기석 목사, 67세, 남성

https://youtu.be/_TZczyVW4vw?si=hkUpLYJ4IsyzH1cz

훈련 데이터와 같은 방법으로 묵음구간을 선택, 다만 노인과 청년의 길이를 맞춰야했던 (확률문제 때문에) 문제가 없기때문에 전체 부분에 대해서 테스트를함.

folder : test/1~5 의 5개의 음원 파일

준비 자체는 음성파일만 준비했으며, 정규화와 이동평균필터의 경우, 이를 적용해서 음성파일을 따로 저장하진 않고, 처리된 결과값을 벡터화 시켜서 사용했습니다.

Constructure

음성 데이터

eLDA 보고서에서의 전처리는 다음과 같이 이루어졌습니다.

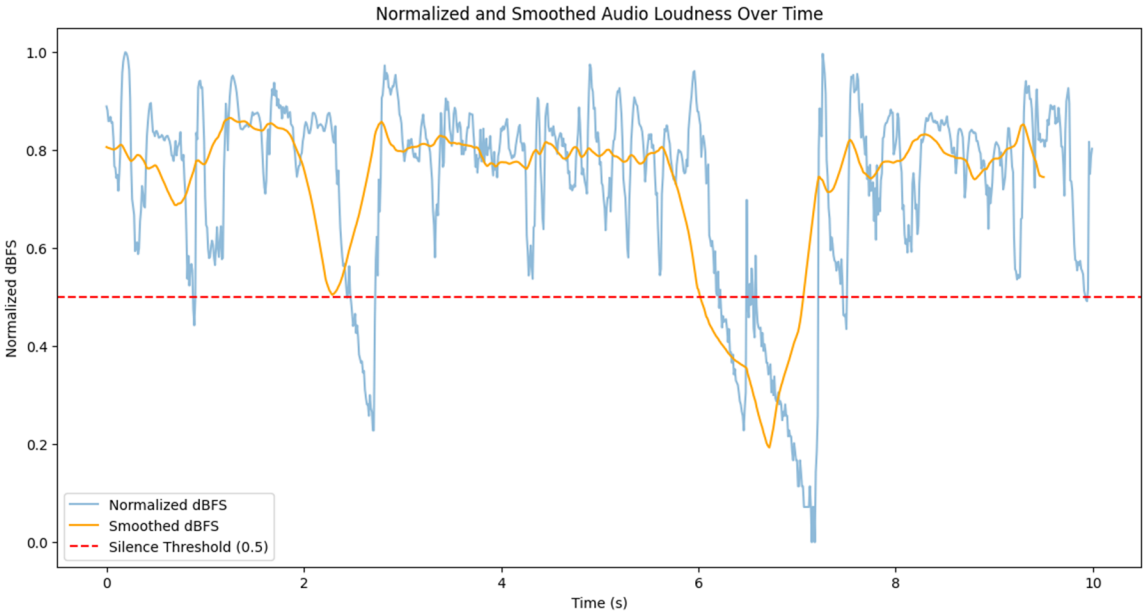
1. **정규화** : 녹음 환경이 전부 제각각임으로, 편의상 1~0으로 정규화시켜 환경을 통일시키고 해석 편의성을 높임
   참고 : 완전 무음 구간에 대해 데시벨 값이 -inf가 나옴으로 -100db로 지정함. (zero divide 문제가 생김)

2. **이동 평균 필터** : 인접한 n개의 데이터 평균을 구하여 순차적으로 데이터를 필터링 하는 기법으로.

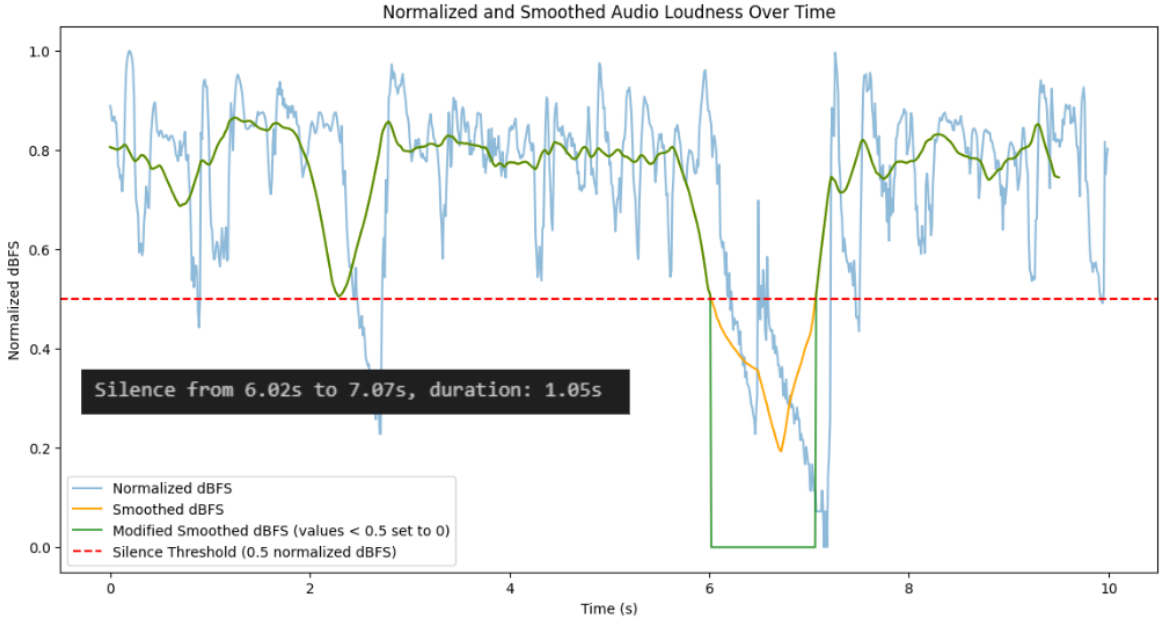
   변동성이 심한 데시벨 영역에서 일정 데시벨 이하를 무음 구간을 잡게 될 경우, 유의미한 분석이 불가능해짐.

   이에 그래프 추이를 보여주는 이동 평균 필터를 사용해 그래프의 추이를 구했습니다.

본 보고서에서 주변 데이터 50개의 평균을 냄. 그 그림은 다음과 같음.



위 사진은 음성 파일중 10초를 잘라 온것. 파란색이 원본 데시벨. 노란색이 이동 평균



초록색인 부분이 threshold(0.5) 이하로 내려갈 경우 0으로 만들어 0인 구간에 대해서 기록하고, 다시 0이 아니게 되는 시점을 기록해 무음 구간을 구함.

Bayes' theorem

새로운 데이터 x가 들어왔을 때, 어떤 클래스에 속할 확률이 더 높은지 확인하는 과정이 확률적 LDA의 핵심이다.

$$P(C = \text{노인} | X = x)$$

따라서, 위와 같이 새로운 데이터 x가 들어왔을 때, 노인 class에 속할 사후확률을 계산하면 된다.

이 사후확률을 베이즈 정리로 나타내 보려고 한다.

$$= \frac{P(X = x | C = \text{노인}) P(C = \text{노인})}{P(X = x)}$$

$$= \frac{f_{\text{노인}}(x) \pi_{\text{노인}}}{\sum_{i=1}^{2} f_i(x) \pi_i}$$

여기서 우리는 노인 class의 데이터 확률밀도 함수인 $f_{\text{노인}}(x)$를 지수분포를 사용한다.

[지수 분포 (Exponential distribution)]

확률변수 X의 확률밀도함수가

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & 0 < X < \infty \\ 0, & \text{기타} \end{cases}$$

$$\mu = E(X) = \frac{1}{\lambda}$$

$$\sigma^2 = V(X) = \frac{1}{\lambda^2}$$

지수분포를 도입한 이유는 다음과 같습니다.

지수분포는 첫 번째 사건이 일어나기까지 대기시간에 대한 분포를 나타낸다.

우리는 화자가 말하기 까지 대기 시간인 묵음기간의 길이로 확률분포를 구성한다.

이와 같이 eLDA는 음성 데이터의 시간적 연관성을 강화하기 위해 지수 분포를 활용한다.

$$P(X = x | C = \text{노인}) = f_{\text{노인}}(x) = f(x; \lambda_{\text{노인}}) = \lambda_{\text{노인}} e^{-\lambda_{\text{노인}} x}$$

우리는 이 '대기시간'에 초점을 맞추어 말하기까지 대기시간 즉, 무음기간의 길이로 노인과 청년 화자를 비교하려는 것이다.

$$\lambda_{\text{노인}} = \frac{N}{\sum_{i=1}^{N} w_i x_i}$$
$$w_i, x_i \in \text{노인}$$

우리는 기존의 지수분포에 각 데이터 x에 대해 가중치 w를 도입하여 람다값을 재정의하였다.

w를 도입하는 이유는 2가지이다.

1. 평균에서 많이 떨어진 이상치 데이터는 가중치를 적게 하여 평균의 응집성을 높인다.
2. 평균의 응집성을 높이기 위해 학습할 parameter를 가중치 w로 하여 w에 대해 목적함수를 편미분한다.

EM 과정

$$\lambda' = \frac{N}{\sum_{i=1}^{N} w_i x_i}$$

w를 도입하여 새롭게 재정의한 λ'이다.

이렇게 재정의한 λ'를 EM과정의 Expectaion 단계의 responsive 변수로 할당한다.

$$f(x; \lambda') = \lambda' e^{-\lambda' x}$$

responsive 변수인 λ'를 도입하여 새롭게 확률밀도 함수인 지수분포를 재정의하였다.

$$\ln L(\lambda') = \sum_{i=1}^{N} (\ln \lambda' - \lambda' x_i)$$

재정의한 지수분포로 log likelihood를 재정의 하였다.

이 재정의된 log likelihood가 우리 모델의 목적함수가 된다.

log likelihood는 모델이 관측된 데이터를 얼마나 잘 설명하는지를 나타내는 지표이다. 이 값이 클수록 모델이 데이터를 잘 설명한다는 것을 의미한다.

따라서 우리는 최적화의 방향을 log likelihood를 최대화 하는 방향으로 학습한다.

이때, 만약 w를 도입하지 않았다면

$$\frac{d}{d\lambda}\ln L(\lambda) = \frac{N}{\lambda} - \sum x_i = 0$$

학습할 parameter가 λ 이므로, 이에 대해서 미분을 진행하고 0이되는 학습된 λ를 구해보면

$$\lambda_{new} = \frac{N}{\sum x_i}$$

학습된 λ_{new}는 평균의 역수인 기존 λ와 동일한 것을 볼 수 있다.

이는 새로운 데이터가 들어오지 않는 이상 평균은 달라지지 않으므로 λ값은 변하지 않게 된다.

$$w' = w + \alpha\frac{d}{dw}lnL(\lambda)$$

따라서, 우리는 w에 대해 목적함수인 log likelihood를 미분하여 목적함수를 최적화 하는데 기인하는 w'값으로 한다.

$$\lambda_{new} = \frac{N}{\sum w_i' x_i}$$

최적화된 w'를 통해 람다값을 재정의 한다.

위와 같은 과정을 통해 실제로 w를 학습하게 되면, 이상치에 대한 가중치 값만 줄어들어 좀 더 평균에 응집하게 된다.

```
x = np.array([0.6, 0.4, 5.6, 0.5, 0.7])# 데이터
```

실제로, 데이터가 위와 같이 5.6만 평균에서 떨어진 이상치라고 가정했을 때,

```
iteration 0: [0.99 0.99 0.99 0.99 0.99]
iteration 10: [0.98606596 0.98737731 0.95328232 0.98672164 0.98541029]
iteration 20: [0.98201687 0.98467791 0.91549075 0.98334739 0.98068634]
iteration 30: [0.97784199 0.98189466 0.87652526 0.97986833 0.97581566]
iteration 40: [0.97352885 0.97901923 0.83626929 0.97627404 0.97078366]
iteration 50: [0.96906274 0.97604183 0.79458561 0.97255229 0.9655732 ]
iteration 60: [0.96442616 0.97295077 0.75131081 0.96868847 0.96016385]
iteration 70: [0.95959797 0.96973198 0.70624775 0.96466498 0.95453097]
iteration 80: [0.95455231 0.96636821 0.65915492 0.96046026 0.94864437]
iteration 90: [0.9492569  0.96283793 0.60973105 0.95604742 0.94246638]
Optimal weights: [0.94424372 0.95949581 0.56294137 0.95186977 0.93661767]
```

학습을 진행하면, 5.6에 대한 w값이 훨씬 낮아져 평균에 더 응집하는 효과를 얻는다.

경사상승법

$$w' = w + \alpha \frac{d}{dw} lnL(\lambda)$$

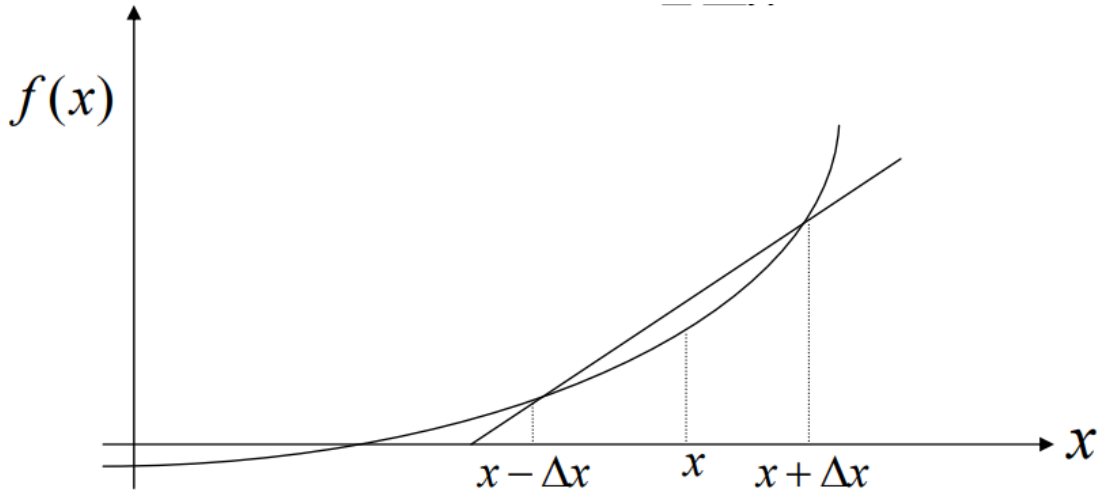
결론적으로 우리는 log likelihood 목적함수를 w에 대해 미분한다.

이 최적화 방향은 log likelihood를 최대화하는 방향이므로, 경사상승법을 진행한다.

하지만, 미분식에 log가 들어가 있는 것을 볼 수 있다.

우리는 기계학습 수업시간을 통해 log가 들어가 있으면 미분이 복잡하여 다른 방법으로 접근해야 한다고 알고 있다.

$$f'(x) \cong \frac{f(x+\Delta x) - f(x-\Delta x)}{2\Delta x}$$



따라서, 우리는 미분값의 근삿값을 구하기 위해

중앙차분법을 이용한다. 중앙차분 법이란, x보다 조금 큰 지점과 조금 작은 지점의 평균기울기를 이용해

미분값을 근사하는 방식이다. 우리는 log likelihood를 w에 대해 미분하는 과정을 중앙차분법을 이용하였다.

우리는 학습이 중단되는 조건을 우리의 목적함수 값인 log likelihood의 변화가 일정 수준(0.001)이하면 학습이 종료되게 끔 설정하였다.

$$|\ln L(\lambda_{new}) - \ln L(\lambda)| <= 0.001$$

선형판별식

새로운 데이터 x의 무음구간 $\{x_1, x_2, ..., x_i\}$

학습이 완료된 노인과 청년의 확률밀도 함수값을 기반으로 새로운 데이터 x의 class를 예측하는 선형판별식을 세워본다.

$$a = ln(\frac{P(C = \text{노인}|X = x)}{P(C = \text{청년}|X = x)})$$

데이터가 클래스에 속할 사후확률을 로짓변환을 거쳐서 위와 같이 선형판별식 a를 세웠다.

이 a는 새로운 데이터 x가 노인 class에 속한다면 a > 0 이고,

새로운 데이터 x가 청년 class에 속한다면 a < 0이다.

$$= ln(\frac{P(X = x|C = \text{노인})P(C = \text{노인})}{P(X = x|C = \text{청년})P(C = \text{청년})})$$

이제 이 식을 좀 더 정리해 보겠다.

$$P(C = \text{노인}) = P(C = \text{청년}) = 0.5$$

여기서 우리는 노인과 청년 데이터를 동일하게 준비하여 균등사전확률을 따르도록 하였다.

$$a = ln(\frac{f_{\text{노인}}(x)\pi_{\text{노인}}}{f_{\text{청년}}(x)\pi_{\text{청년}}})$$

$$= \ln \frac{\lambda_{\text{노인}} e^{-\lambda_{\text{노인}} x}\pi_{\text{노인}}}{\lambda_{\text{청년}} e^{-\lambda_{\text{청년}} x}\pi_{\text{청년}}}$$

$$= \ln \frac{\lambda_{\text{노인}}}{\lambda_{\text{청년}}} - \lambda_{\text{노인}} x + \lambda_{\text{청년}} x + \ln \frac{\pi_{\text{노인}}}{\pi_{\text{청년}}}$$

$$a = x(\lambda_{\text{청년}} - \lambda_{\text{노인}}) + \ln \frac{\lambda_{\text{노인}}}{\lambda_{\text{청년}}} + \ln \frac{\pi_{\text{노인}}}{\pi_{\text{청년}}}$$

위와 같이 식을 정리하여 선형판별식 a를 얻었다.

이제 이 a값을 변형하여 sigmoid(a) 형태로 바꾸어 보겠다.
sigmoid(a) 형태로 바꾼다면, 출력값이 0~1 사이이며, 이 출력값이 0.5보다 크다면 노인에 속하며, 0.5보다 작다면 청년에 속하도록 좀 더 확률적으로 접근하기 용이해진다.

$$e^a = \frac{P(X = x|C = \text{노인})P(C = \text{노인})}{P(X = x|C = \text{청년})P(C = \text{청년})}$$

$$1 + e^{-a} = \frac{P(X = x|C = \text{청년})P(C = \text{청년}) + P(X = x|C = \text{노인})P(C = \text{노인})}{P(X = x|C = \text{노인})P(C = \text{노인})}$$

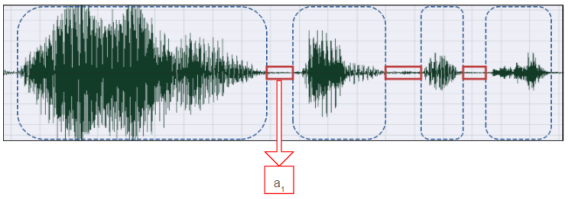
$$= \frac{P(X = x|C = \text{노인})P(C = \text{노인})}{P(X = x)}$$
$$= sigmoid(a)$$

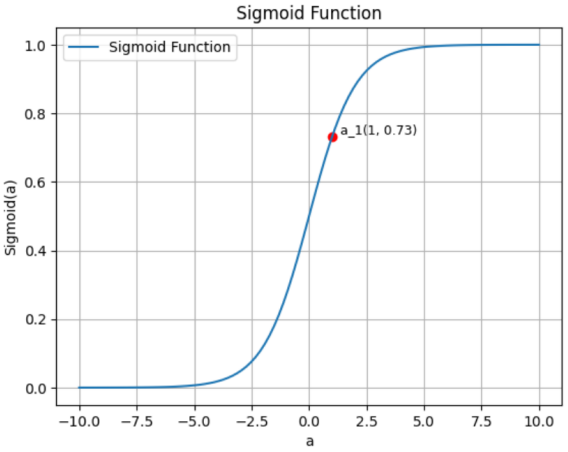
$$\frac{1}{1 + e^{-a}} = \frac{P(X = x|C = \text{노인})P(C = \text{노인})}{P(X = x|C = \text{청년})P(C = \text{청년}) + P(X = x|C = \text{노인})P(C = \text{노인})}$$

새로운 데이터x를 넣었을 때,
시그모이드 함수값 > 0.5 이면, 노인 class에 속한다.

시그모이드 함수값 < 0.5 이면, 청년 class에 속한다.

Prediction



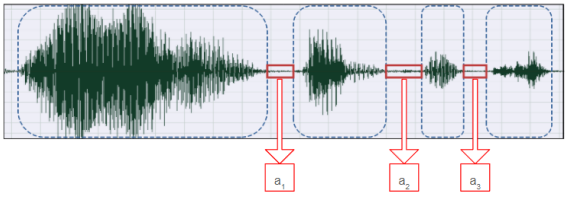


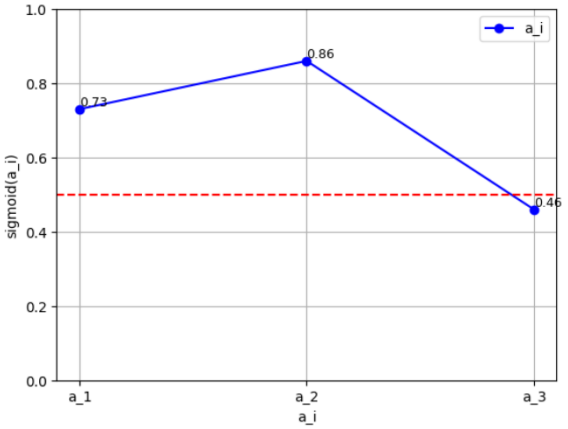
만약 데이터 x의 첫 번째 무음구간 x_1의 선형판별식이 a_1 = 1이라고 한다면,

sigmoid(1) = 0.73이다. 즉, 이 데이터 x_1의 무음구간에 대해서는 73%의 확률로 노인일 것이다.

실제 데이터는 $x_1, \ldots, x_n$같이 연속적인 데이터이다.

실제 시간이 바뀜에 따라 데이터가 노인일 확률이 변할 것이다.





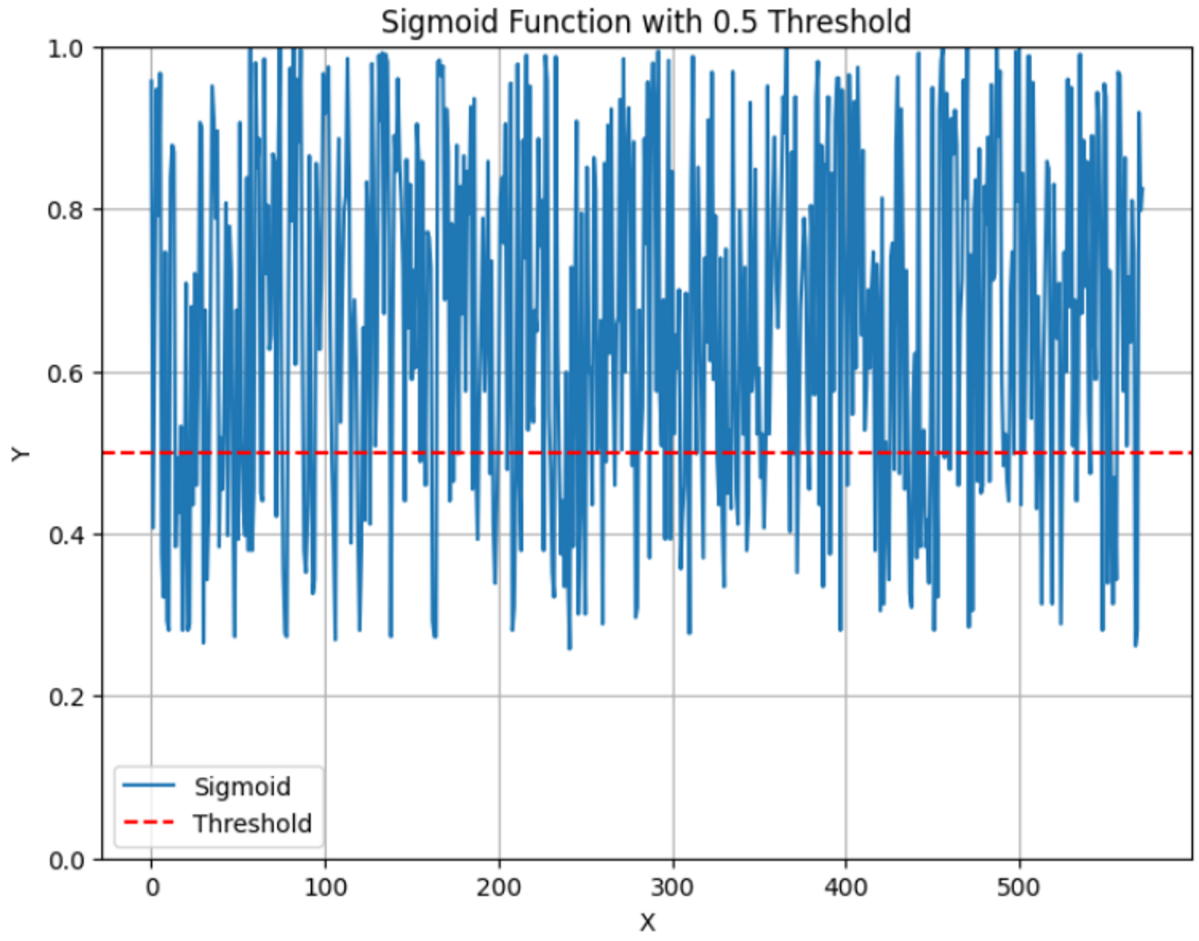
sigmoid(a1) = 0.73
sigmoid(a2) = 0.86
sigmoid(a3) = 0.46

a에 대한 sigmoid 출력값이 위와 같다고 가정한다면,

데이터는 시간이 지날수록 노인일 확률이 73% → 86% → 46%로 바뀌게 될 것이다.

그리고 평균값이 0.68이므로 노인일 확률의 평균은 68%이다.

실제 결과값



김기석 목사(67세) 테스트 데이터에 대한 예측 결과이다.

선형판별식에 sigmoid function을 도입해 확률값을 0~1로 하였다.

시각화한 결과 위와 같이 0.5임계값을 넘는 경우가 많았고, 평균은 65.9%이다.

즉, 테스트 데이터가 노인 class에 속할 확률은 평균적으로 65.9%이다.

```
테스트 데이터가 노인일 확률(평균) : 65.97667744082429%
```

기존 LDA와 비교

기존 LDA는 우리 모델과 다르게 각 class에 속할 확률을 제공하지 않는다.

따라서 우리의 모델도 sigmoid 값이 0.5보다 크면 1 아니면, 0으로 하여 전체 값을 평균내서 좀 더 hard하게 예측하였다.

동일한 방법으로 예측을 진행하였을 때,

정확도가 우리의 모델이 13%정도 더 높게 나왔다.

이는 당연한 결과이다. 우리의 모델은 무음구간에 대한 데이터 즉, 대기시간에 맞는 확률분포인 지수분포를 사용하였고,

기존 LDA는 이상치에 대해 민감하다는 약점이 있다. 하지만, 우리의 모델은 이상치에 대한 민감도를 낮추기 위해 w를 도입하여서

이상치에 대한 민감도를 대폭 낮추었다.

그리고 기존 LDA는 가우시안 분포를 사용하여 데이터가 많아질수록 공분산행렬에 대한 연산량이 cost가 높지만,

우리의 모델은 지수분포를 써서 평균을 내는 방식이기에 훨씬 compact하다.

Conclusion

본 연구에서는 기존 선형 판별 분석(LDA)이 동적이고 복잡한 음성 데이터를 처리하는 데 있어 여러 가지 한계를 지니고 있음을 지적하였습니다.

특히, LDA는 음성 데이터의 변동성과 복잡성을 효과적으로 반영하지 못해 실용적인 응용 사례가 제한적이었습니다.

이러한 문제를 해결하기 위해 본 연구에서는 지수 분포 기반의 새로운 LDA 변형 기법인 지수형 선형 판별 분석(eLDA)을 제안하였습니다.

eLDA는 지수 분포를 도입하여 음성 데이터의 시간적 연관성을 강화하였으며, 가중치 시스템을 통해 이상치에 대한 저항성을 향상시켰습니다.

이를 통해 음성 데이터의 특성을 보다 효과적으로 반영할 수 있게 되었습니다.

본 연구에서는 기존 LDA와 eLDA를 동일한 데이터 셋에 적용하여 정확도를 비교 분석하였습니다.

실험 결과, eLDA는 기존 LDA보다 음성 데이터 분석에서 더 높은 정확도를 보였습니다. 구체적으로, eLDA는 기존 LDA보다 약 13% 더 높은 정확도를 기록하였습니다.

이는 eLDA가 음성 데이터의 특성을 보다 정밀하게 반영하며, 이상치에 대한 민감도를 낮춤으로써 보다 안정적이고 신뢰성 있는 분석을 가능하게 한다는 점을 시사합니다.

또한, eLDA는 공분산 행렬의 계산 복잡성을 줄여 더 효율적인 연산이 가능하게 하였습니다. 이는 음성 데이터의 분석에서 eLDA가 기존 LDA에 비해 계산 비용 측면에서도 우수함을 보여줍니다.

향후 연구에서는 더 다양한 데이터셋과 여러 가중치 요소들을 추가하여 eLDA의 성능을 더욱 향상시키는 방향으로 진행될 필요가 있습니다.

이를 통해 eLDA가 음성 데이터 분석의 다양한 응용 분야에서 더욱 광범위하게 사용될 수 있을 것으로 기대됩니다.

결론적으로, 본 연구는 지수 분포 기반의 eLDA가 기존 LDA에 비해 음성 데이터 분석에서 더 높은 정확도와 효율성을 제공함을 입증하였습니다. 이를 통해 eLDA는 음성 데이터 분석의 새로운 대안으로서 그 잠재력을 확인하였습니다.