Q. 공채와 관련하여 포트폴리오는 어떻게 준비해야 하는지 궁금합니다.

A. 신입공채는 특별한 커리어 보다는 자세(Attitude)를 보려는 경향이 있습니다. 열정과 하려는 자세가최대한 드러나는 것을 어필하는 것이 좋습니다.

<빅데이터 분야 석박사 학위자 입사시험문제>

주어진 가상의 데이터셋(테이블 형태) 을 주어주고 어떻게 중요 feature를 도출하고 분석결과를 도출할 것인지 가정을 설명하라는 문제가 주어짐(데이터 해석, 전처리방법, 스코어링 방법, 가설 수립 등에 제한이 없고, 자유롭게 논리적으로 설명할 것이라고 명시)

<빅데이터 프로세스 분석 과정>

* 분석 프로세스

주제 선정 -> 데이터 수집 -> 전처리 및 저장 -> 데이터분석/모델링 -> 모델검증

최적화 ↖ ← |

주제 범위 선정 -> 세부 요구사항 수립 -> 필요 데이터 탐색 -> 데이터 수집 방법 결정 ->

데이터 수집 / 저장 -> 데이터 품질 확인 -> 데이터 마트 생성 -> 성능검증(A/B Test) ->

모델 검증 -> 데이터분석/ 모델링 -> 전처리 -> 데이터 탐색 -> 운영 시스템 적용 ->

모니터링 -> 유지보수

☞주제를 선정한 프로젝트의 예시

주제 선정: 주제(고양이 맞추기)

세부 요구사항

1. Input : 고양이 그림 Output: 고양이
2. 데이터 수집: Crawling & Download

데이터 수집:

전처리

분석/모델링

모델 검증

< 인물로 알아보는 NN 역사>

-----|---------------------|------------------------|-----------------------|---------------------|-------------------|--------🡪

1943 1957 1969 1974 1986 2006

제프리 힌튼

Deep Belief

Network

(심층신뢰신경망)

이라는

계층간에는 연결이 있지만

걔층 내의 유닛

간에는 연결이

없는 신경망 제시

↓

Vanishing

Gradiant

문제 해결

제프리 힌튼,

역전파를 재발견하여

세상에 내놓음

Multi-layer

Network

(Deep neural

Network)

를 제시함

힌튼,

데이비드

럼멜하트,

로널드

윌리엄스가

오류 역전파

알고리즘으로

다중 퍼셉트론을

학습시키는데

성공함

폴 워보즈가

박사과정 논문에서

‘역전파

(Backpropagation)’

를 제안

앞의 진행방향에서

고쳐가는 것이 아닌

결과를 보고 뒤로

가면서 weight와

Bias를 조정하는

알고리즘

마빈 민스키,

시모어 패퍼트가

퍼셉트론의 한계를 수학적으로 증명

(XOR연산이 불가능)

헵 의 학습이론

(두 뉴런 사이의 연결강도를 조정할 수 있는 학습이론)

“Cell A가 B를 자극시키는데 충분하고, 지속적인 원인으로 작용한다면, cell B에 대한 cell A 작용력이 내부적으로 변화를 일으켜 증가하게 된다”

Ex/ 파블로프의 개 실험(신경강화)

워터필츠,

워렌맥컬린

프랭크 로젠블럿의

‘퍼셉트론’

최초의 신경망

모델을 발표

선형분류를

수행할 수 있는

‘피드포워드

뉴럴네트워크’

제안

<Backpropagation 의 이해>

forward propagation을 통해 필요한 값들을 미리 저장해두고, backward propagation이 진행되면서 위에서부터 loss에 대한 derivative를 하나하나 계산해나가면서 다음 layer에서 바로 전 layer에서 계산한 값들과 각 neuron 별로 추가적으로 필요한 derivative들을 곱해나가면서 weight의 derivative를 계산하는 알고리즘이다.

이렇게 한 번 전체 gradient를 계산한 다음에는 learning rate를 곱하여 전체 parameter의 값을 update한 다음, 다시 처음부터 이 과정을 반복한다. 보통 에러가 감소하는 속도를 관측하면서 ‘이 정도면 converge한 것 같다’ 하는 수준까지 돌린다.

익숙해지려면 다소 시간이 걸리지만, 개념적으로 먼저 ‘error를 먼저 계산하고, 그 값을 아래로 전달해나가면서 바로 전 layer에서 계산한 미분값들을 사용해 현재 layer의 미분값을 계산한 다음, 그 값을 사용해 다음 layer의 미분값을 계산한다.’ 라고 개념만 이해해두고 다시 차근차근 chain rule을 계산해나가면서 계산하면 조금 편하게 익숙해 질 수 있을 것이다.