**Machine Learning Theory & Foundation**

**Quiz no.1**

**Q1**. Tom Mitchell 의 Machine Learning 정의에는 ETP모델이라는 개념이 나옵니다.  
스팸 분류 서비스를 만드려고 할 때 다음 예시들을 보고 E,T,P로 구분하시오.

스팸인 메일과 스팸이 아닌 메일을 분류한다. ( T )  
이용자가 스팸이라고 분류한것과 일반 메일을 관찰한다. ( E )  
스팸여부를 정확히 분류한 메일의 개수를 측정한다. ( P )

**Q2**. 머신러닝의 종류 2가지를 서술하고 각각의 종류에는 어떤 머신러닝 기법이 있는지 쓰시오.

(종류1 : 지도 학습(Supervised Learning) , 해당 기법 : Regression,Classification)  
(종류2 : 비지도 학습(unsupervised Learning) ,해당 기법 : Clustering)

**Q3**. What is the algorithm to predict continuous valued output such as price?   
( Regression )

**Q4**. What is the algorithm to predict Discrete valued output such as Boolean variables?  
( Classification )

**Q5**. 상황에 맞는 머신러닝 기법을 분류하시오.  
상황 1 : 용구는 유명한 약장수이다. 용구의 말주변은 매우 뛰어나서, 실적이 매우 좋다. 여태까지의 용구의 약 판매 수량을 이용하여 3개월동안 용구가 팔 수 있는 약의 개수를 예측하려고 한다.

이 때 사용할 머신러닝 기법은?

( Regression )

상황2 : 구익이는 ROTC로 활동하고있다. 어느날 17학번 후배가 ROTC를 지원하고 싶은데 자신이 합격을 할 수 있을지 물어보았다. 구익이는 여태까지 ROTC를 합격한 사람들의 사례를 통해 머신러닝을 이용하여 후배의 합격 가능 유무를 알아보려고 한다. 이 때 사용할 머신러닝 기법은?

( Classification )

**Q6**. 비지도학습(Unsupervised Learning)은 무엇인지 서술하시오.

( Labeling된 값 없이, 답이 정해지지 않은 데이터를 제공하여, 데이터들의 동일한 구조상의 특징을 찾아 내는 것이다. )

**Q7**. 구글 뉴스의 비슷한 기사 모아보기, 페이스북의 친구 추천, 시장 고객 분석등에 사용되는 기법은? ( Clustering )

**Q8**. 비지도 학습을 사용해야 하는 경우를 모두 고르세요. ( 2,3,5 )

1. 스팸여부에 따라 분류된 메일들을 통해 Spam Filter를 학습시킨다.
2. 웹에서 비슷한 내용을 다루는 뉴스기사들을 묶어서 보여준다.
3. Customer데이터를 통해 자동적으로 고객들을 어떤 상품을 선호하는지에 따라 다양한 소비군으로 분류한다.
4. 이미 진단이 완료된 환자들의 종양데이터를 통해, 새로운 환자의 종양이 악성인지 양성인지 분류한다.
5. 시끄러운 음악이 흘러나오고 사람들의 말소리가 가득한 파티현장에서 사람의 목소리만을 추출하거나, 음악만을 추출한다.

**Q9**. What is Hypothesis in machine learning?  
( 우리가 예측하고자 하는 값에 대한 식으로 weight\*Feature꼴로 나타난다. )

**Q10**. What is Cost Function in machine learning?  
( Cost Function이란, 예측한 Hypothesis값과 실측값 y의 차이를 나타내는 지표이다. )

**Q11**. If we use Linear Regression, What is the formula which calculate   
( )

**Q12**. Learning Rate를 이용하여 가장 오차가 작아지는 Theta값을 찾는 방법은 무엇인가?  
( Gradient Descent )

**Q13**. Learning Rate 값이 너무 크거나 작을 경우 생길 수 있는 문제점은 무엇인가?  
의 값이 너무 큰 경우 : ( Cost Function의 값이 Iteration이 늘어도 줄지 않는다. )  
의 값이 너무 작은 경우 : ( Cost Function의 값이 너무 천천히 줄어든다. )

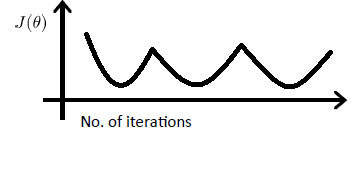
**Q14**. Feature Scaling을 하는 이유는 무엇인가?

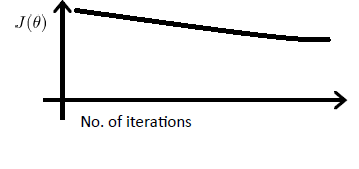
( 값들의 범위를 동일하게 설정하여 Gradient Descent를 보다 빠르게 하기 위해서 )

**Q15**. 다음 값들을 Mean Normalization 방법으로 Feature Scaling 하시오.  
집의 크기()의 범위 0  
화장실의 개수의 범위 0 < x2 < 5

-0.5 < x1 < 0.5, -0.4 < x2 < 0.6

**Q16**. 다음 그래프는 Gradient Descent를 진행에 따른 Cost Function의 변화이다. 이 그래프의 모모양을 봤을 때, Learning Rate를 어떻게 수정해야 하는가?

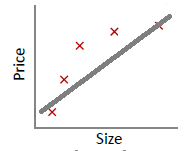
 ( Learning Rate의 값을 줄인다.)

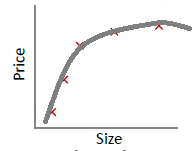
 ( Learning Rate 값을 늘린다.)

**Q17**. Weight를 반복없이 한번에 구할 수 있지만, Feature종류가 많아질수록 느려지는 알고리즘은 무엇인가? ( Normal Equation )

**Q18**. Logistic Regression 에서 Hypothesis로 사용하는 함수의 이름과 식을 작성하시오.  
 ( Sigmoid(Logistic) Function , )

**Q19**. Logistic Regression 에서 y = 1 일 확률이 30%일 경우 h(x)의 값을 구하시오. ( 0.3 )

**Q20**. 그래프를 보고 알맞은 단어를 고르시오.  
(Over fitting, **Under fitting**) (**high bias**, high variance)

(**Over fitting**, Under fitting) (high bias, **high variance**)

**Q21**. Overfitting의 해결 방법은 무엇인가?  
( Regularization을 이용하거나, Feature의 개수를 줄인다. )

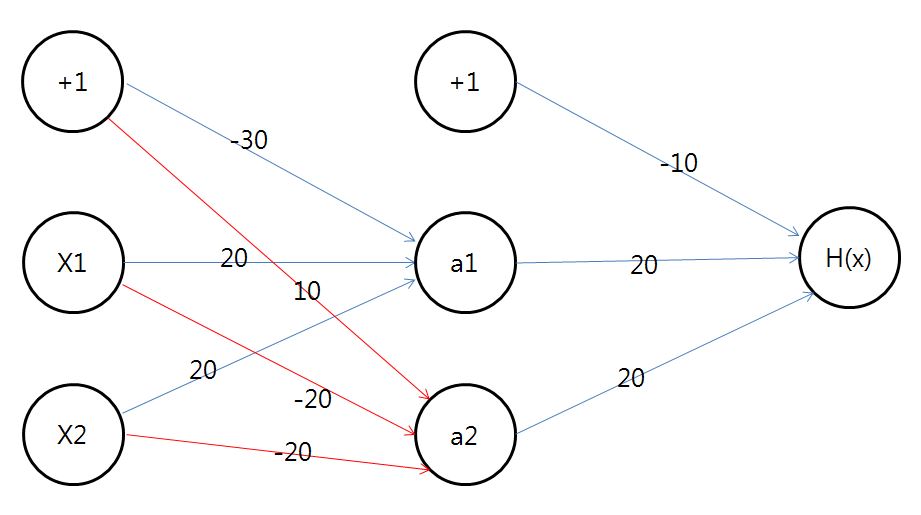
**Q22**.   
Regularization Feature 가 너무 클 경우 어떤 현상이 발생하는가? ( High bias화 된다. )  
Regularization Feature 가 너무 작을 경우 어떤 현상이 발생하는가?( High Variance화 된다. )

**Q23**. 딥러닝이 무엇인가?

( Deep Neural Structure(Hidden Layer가 2개이상)을 이용한 기계 학습방법의 일종이다. )

**Q24**. Layer의 종류 세가지를 서술하시오.

( Input Layer ), ( Hidden Layer ), ( Output Layer )

**Q25**. Neural Network 개념을 이용하여, XNOR을 구현하시오.  


**Q26**. Backpropagation Algorithm의 목적은?

( Gradient Desecent에 필요한 Cost Function의 편미분값을 구하기 위해서 )

**Q27**. Gradient Checking이 무엇인가?

( 수학적으로 미분계수를 직접 계산하고, Backpropagation과 비교하여 정상 작동하는지를 체크 )

**Q28**. Neural Network의 학습 순서대로 숫자를 채워넣으시오.  
Cost Function을 계산하는 코드를 작성한다. ( 3 )  
Gradient Descent를 통해 Weight를 구한다. ( 6 )  
Back Propagation을 통해 를 구한다. ( 4 )  
Weight를 Random하게 Initialize한다. ( 1 )  
Gradient Checking을 통해 Back Propagation을 검사한다. ( 5 )  
Forward Propagation을 통해 h(x) 값을 구한다. ( 2 )

**Q29**. Cross Validation을 이용하여 ( Data Model(차수) ) 을 먼저 구한뒤, Train set을 이용하여 Weight를 학습시킨다.

**Q30**. Tensorflow의 강력한 기능중 하나로써, 변수의 값을 넣지는 않고, 변수의 데이터형만을 선언하여, 가변적인 연산을 할 수 있도록 하는 도구는 무엇인가? ( Placeholder )