

# 2022 Artificial Intelligence: Final Exam Essay

ID and name: 20186889 권용한

You are a manager at "Awesome Homes," a real estate company specializing in vacation home rentals. Over the last five years, sales have decreased 10% annually, and customer feedback indicates that your company's rental prices are seen as too low or too high relative to market value.

The rental prices of Awesome Homes are set through traditional means in which a team of real estate agents monitor various market metrics, and use their expertise to determine rental value. Your CEO has tasked you with managing a new team which will employ machine learning techniques in order to improve the rental price estimates.

1. Many of the real estate agents at Awesome Homes, close personal friends of yours, are worried about being replaced by computers. How can you gain their buy-in on this new initiative?

인공지능을 제작함에 있어서 분야의 전문가의 도움 및 의견 필요성을 알리고, 제작된 인공지능 만큼 동등한 것으로 보는 것이 아니라, 가격 정책을 돕는 역할만을 분명히 하여 인공지능이 그들을 대체하는 것을 막을 수 있습니다.

2. Is this initiative a supervised or unsupervised problem? Explain your reasoning.

Supervised 입니다. 그동안의 데이터를 통해 가격을 예측하는 것이므로, 구하고자 하는 바가 명확한 supervised 입니다.

3. What types of data do you need to gather in order to proceed?

지역적 특징 (지역, 고도, 평면 기온, 강수량, 호수...), 편의 시설 (학교, 병원, 마트, 여가시설을 즐길 수 있는 공간), 이용 가능 교통편과 거리, 치안, rental 기간 (시작일, 종 rental 기간)

Your new team presents several variations of machine learning models for estimating rental prices.

4. How do you train and test each model?

train과 test 데이터는 구분이 필요하므로 전체 데이터 중 20%의 데이터를 test 데이터로 사용합니다.

train 데이터 중 1%의 비율로 Cross Validation을 진행하여, 이를 통해 더 정확도 높은 모델을 생성합니다.

5. By what metrics will you determine the quality of the models' performance?

실행을 통해 예측되는 비용과 실제 비용의 오차를 줄이는 것이 우리의 목표입니다.

따라서 성능은 Mean Squared Error (MSE)로 측정합니다.

6. How will you determine whether your models are coming up with better rental price estimates than were obtained from the traditional reliance on real estate agents' expertise?

같은 실험 데이터를 전통적 방식으로 예측한 결과와, 머신러닝을 사용한 결과, 그리고 실제 비용 총 3가지 데이터를 측정하고 MSE를 측정해 전통적 방식과 머신러닝의 정확도를 수치적으로 비교할 것입니다.

You applied one of your models to the training data, the model's predictions of rental prices is sound and consistent. However, the model's success is seriously compromised when new unseen data are entered.

7. What is going on with this model?

train 데이터에 값을 너무 맞추어 overfitting이 발생되었다고 생각합니다.

8. What steps can be taken to improve this model's prediction capabilities?

overfitting을 줄이기 위해 Regularization이 필요합니다. 예측 결과가 continuous 한 것이므로 머신러닝 모델을 regression을 사용할 것이며, 따라서 Ridge, Lasso, Elastic Net Regularization 중 모델에 적합해 보이는 Regularization과 parameter를 조정해 보려는 재조정 합니다.

Your team has assembled three models whose success rates are more or less identical. Here are some observations:

Model	Time for train	Speed of prediction	Average prediction error
A	24 hours	One second or less	5%
B	Six hours	Six hours	1%
C	Five minutes	Two minutes	10%

9. Which model will you recommend to your CEO? Explain your decision.

예측이 6시간이 소요되는 B는 실제 상황에서 사용하기 힘들 것으로 생각되어 배제합니다. 예측 모델이 짧은 시간 내에 결과가 필요한다면, C가 비용적 측면에서 적절하겠지만, 우리는 거의 rental 비용을 예측하므로 한번 생성한 모델을 오랜 기간 사용 가능 할 것으로 생각합니다. (집 값의 변동 기간이 24시간에 비하면 매우 큰 것으로 예상) 또한, 10%의 에러는 사업에 사용하기에 부담스럽게 느껴질 수 있다고 생각해서, A 모델이 가장 적합하다고 생각합니다.