

1. 신경망의 출력이  $(0.4, 2.0, 0.001, 0.32)^T$  일 때 softmax 를 적용한 결과를 쓰시오. (10 pts)

$$e^{0.4} = 1.491 \quad e^{0.001} = 1.001$$

$$e^{2.0} = 7.389 \quad e^{0.32} = 1.378 \quad \therefore \text{softmax} = [0.1725, 0.6562, 0.0889, 0.1224]$$

$$\sum e^x = 11.259$$

2. . 아래 [예제 6-1]에서 초기 군집 중심을  $z_1=x_2, z_2=x_3, z_3=x_7$ 로 다시 설정했을 때, k-평균의 동작을 보이시오. (30 pts)

[예제 6-1]

#### 예제 6-1 k-평균의 동작

[그림 6-5]는 훈련집합이 7개의 샘플을 가진  $n=7$ 인 예를 보여 준다. 좌표는 다음과 같다.

$$x_1 = \begin{pmatrix} 18 \\ 5 \end{pmatrix}, x_2 = \begin{pmatrix} 20 \\ 9 \end{pmatrix}, x_3 = \begin{pmatrix} 20 \\ 14 \end{pmatrix}, x_4 = \begin{pmatrix} 20 \\ 17 \end{pmatrix}, x_5 = \begin{pmatrix} 5 \\ 15 \end{pmatrix}, x_6 = \begin{pmatrix} 9 \\ 15 \end{pmatrix}, x_7 = \begin{pmatrix} 6 \\ 20 \end{pmatrix}$$

군집의 개수  $k=3$ 이라 하자. 맨 왼쪽 그림은 초기 군집 중심을 보여 준다. [알고리즘 6-1]의 라인 3-4는 7개 샘플을 아래와 같이 배정할 것이다.

$$\{x_1\} \text{은 } z_1, \{x_2\} \text{은 } z_2, \{x_3, x_4, x_5, x_6, x_7\} \text{은 } z_3$$

이 배정을 행렬  $A$ 로 표현하면 다음과 같다.

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

[그림 6-5]의 가운데 그림은 새로 계산한 군집 중심이다.  $z_1 = (18, 5)^T, z_2 = (20, 9)^T, z_3 = (12, 16.2)^T$ 이고, 식 (6.2)에 대입하면  $J = 244.80$ 이 된다. 이때 거리함수 dist로 식 (1.7)의 유클리디언 거리를 사용한다.

두 번째 루프를 실행하면 행렬  $A$ 는 아래와 같이 바뀐다. 군집 중심은  $z_1 = (18, 5)^T, z_2 = (20, 13.333)^T, z_3 = (6.667, 16.667)^T$ 이다. 이것을 식 (6.2)에 대입하면  $J = 58.00$ 이 된다. [그림 6-5]의 맨 오른쪽 그림은 두 번째 루프 수행 후의 상황이다.

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}$$

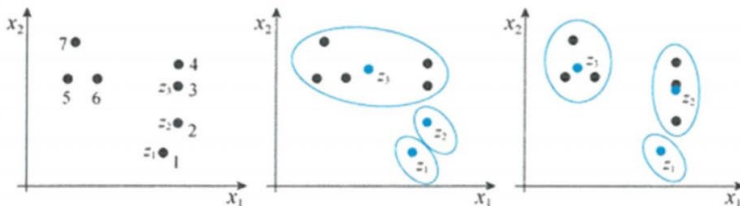
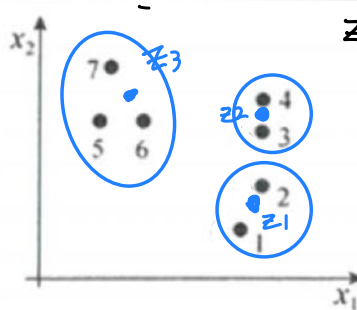
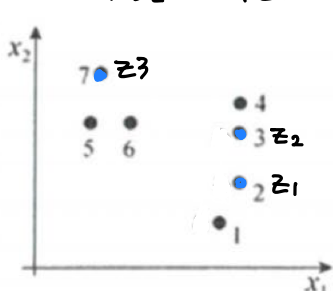


그림 6-5 k-평균의 동작 예제

$$10 + 2.25 + 2.25 + 2.79 + 5.43 + 6.45 + 2.79 + 2.79 + 11.09$$

$$z_1 = x_2, z_2 = x_3, z_3 = x_7$$



$$z_1 = (19, 11)^T \quad z_2 = (20, 15.5)^T \quad z_3 = (6.67, 16.67)^T$$

$$J = 31.84$$

3. [알고리즘 7-1] 은 분류 결과의 신뢰도를 필요로 한다. Softmax 출력을 신뢰도로 간주하는 방안에 대한 견해 (근거, 장점, 단점 및 한계점)를 제시하시오. (15 pts)

Softmax 출력을 신뢰도로 하는 것에 대한 근거는 자가 학습은 신뢰도를 높은 순으로 부분 집합을 선별 해야하는데 Softmax 는 최댓값이 아닌 값을 억제하여 max 함수를 사용하는 모델을 최적화 하기 때문이다.

Softmax 의 장점은 모두 더하면 1 이 되는 확률분포이기 때문에 직관적으로 이해하기가 쉽다.

단점은 과적합의 문제가 생길 수 있으며, softmax 는 다중 클래스 분류 문제를 위해 설계되었기 때문에 다른 유형의 문제에는 적용하기가 쉽지 않다는 한계점을 가지고 있다.

4. 아래 글을 읽고 기술적인 내용에 대해 간략하게 요약하고 이에 관한 본인의 견해 (장점, 단점 및 한계점, 본인이 생각한 개선 방안)를 제시하시오. (45pts)

이 글은 물류 입고 프로세스를 최적화하기 위해 기계학습 모델을 활용한 사례를 설명하고 있다.

풀필먼트 센터로 입고되는 제품의 적정 트럭수를 예측해 운영의 효율성을 높이하고자 한다.

그 과정에서 데이터 수집 및 분석을 거치고 LightGBM 알고리즘으로 모델을 학습하고 베이지안 최적화로 모델의 하이퍼파라미터 탐색을 했다.

한달 주기로 데이터를 업데이트 하여 모델을 업데이트하여 결과적으로 입고 예약 시스템과 모델연계를 했다. 과소 예측의 적절한 균형을 유지하며 과대 예측은 최대한 줄이는 방향으로 최종 모델을 도출했다.

그 결과, 과소 예측 67.9%p 감소, 과대 예측 39.41%p 감소하며 슬롯 운영 효율이 크게 향상 되었다.

이 글을 읽고 기계학습 모델이 현업에 어떻게 적용되는지 지레짐작 이해하는데 큰 도움이 된 것 같다.

데이터 기반 접근으로 보다 객관적이고 정확한 예측이 가능하며 하이퍼파라미터 탐색을 자동화하여 모델 최적화하는 방식은 놀랍다. 실제 업무 시스템과 모델을 연계하여 실용적인 시스템을 만들어 낸다는 것은 기업에 큰 이익을 가져온다. 그러나 모델 예측의 오차로 인한 과소/과대 예측 문제는 여전히 완벽하지 않으며, 새로운 제품군/ 작업 방식 변화에 대한 대응의 한계가 존재한다.

개선방안으로 강화학습으로 접근하는 방법이 있다고 생각한다. 물류 최적화는 순차적인 의사결정으로 접근이 가능하기 때문이다.

또 실제환경과 데이터로만 바라봤을 때의 환경은 어느 정도의 차이가 있을 수 있다고 생각한다. 따라서 실제 환경을 최대한 분석하여 시뮬레이션을 구축하여 다양한 시나리오를 테스트 하는 것도 새로운 전략을 만드는데 있어 효과를 볼 수 있다.

쿠팡은 물류 입고 프로세스의 최적화를 위해 끊임없이 노력하고 있습니다. 풀필먼트 센터로의 제품 입고과정에서 낭비되는 자원을 최소화하면, 적기에 제품을 판매하고 더 많은 고객들에게 더 빠르게 배송할 수 있습니다. 이를 위해 쿠팡은 기계 학습을 통해 직매입 제품들의 풀필먼트 센터 입고 프로세스를 효율적으로 개선해 왔습니다. 본문을 통해 어떤 개선이 있었는지를 살펴보겠습니다. 목차

## · 배경 및 과제

## · 트럭 수량 예측 모델 학습하기

특징 추출 ◦ 모델 학습: LightGBM

## 알고리즘

## ◦ 모델 하이퍼 파라미터 탐색 : 베이지안 최적화

입고 예약 시스템과 모델 연계 구성 ◦ 과소 예측과

과대 예측 사이의 트레이드 오프

## · 모델 적용 결과

## · 향후 계획

## 배경 및 과제

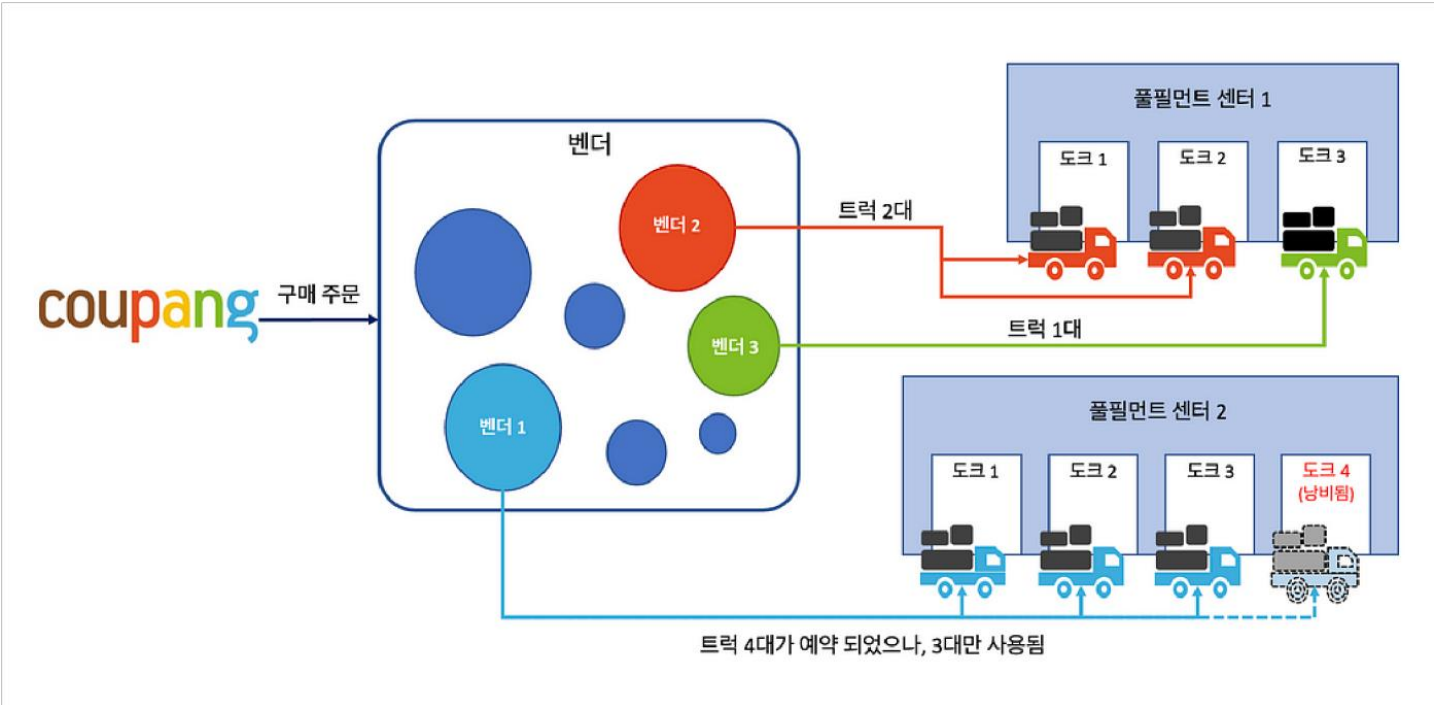
매일 각지의 수많은 업체들이 트럭에 다양한 종류의 제품을 적재해 쿠팡 풀필먼트 센터로 제품을 반입합니다. 각 풀필먼트 센터에는 트럭을 세워두고 물건을 하역하는 도크(dock)가 있습니다. 센터마다 시간대별로 사용 가능한 도크의 최대 개수는 정해져

있습니다. 제품 하역 작업 시 한 대의 트럭이 하나의 도크를 정해진 시간 동안 사용하게 되는데, 이 시간을 슬롯(slot)이라고 부릅니다.

정해진 개수의 슬롯들로 여러 업체들이 납품하는 제품들이 효율적으로 입고되려면 각 입고에 필요한 슬롯 개수가 사전에 정확히 예측되어야 합니다. 만약 사전에 예측한 필요 슬롯 개수가 실제 필요 슬롯보다 적을 경우 입고 과정에서 지연이 발생할 수 있으며, 실제 필요 슬롯보다 많은 경우 한정된 자원인 슬롯을 불필요하게 낭비하게 됩니다.

그림 1. 납품 과정에서 발생할 수 있는 자원 낭비

이에 저희는 업체 특성과 풀필먼트 센터로 입고되는 제품의 특성을 바탕으로, 입고 예약 신청을 하는 업체들에게 적절한 필요 슬롯 개수를 예측해 제공하는 시스템을 개발하기 시작했습니다. 목표는 낭비되는 슬롯의 개수를 줄이고, 슬롯이 부족해



입고가 지연되는 문제를 최소화하는 것이었습니다. 이 목표를 달성하기 위해 저희가 활용한 기술들을 다음 섹션에서 자세히 소개하겠습니다. 트럭 수량 예측 모델 학습하기

저희는 데이터로 문제를 풀고자 했고 이를 위해:

1. 그동안 축적해온 물류 데이터 및 입고 신청 정보로부터 트럭 대수에 영향을 미치는 특징들(feature)을 도출하고, 각 물류 입고에 실제 사용된 트럭 대수 데이터를 결합해 학습 데이터를 구성했습니다.
2. 적정 트럭 입차 수량을 예측하는 머신 러닝(ML) 모델을 학습시켰습니다.
3. 학습된 모델을 입고 예약 시스템과 연결해 업체가 입고 신청을 하면 시스템으로 바로 적정 트럭 대수를 확인할 수 있게끔 설계했습니다.

이렇게 저희는 입고 예약 시스템에 새로운 예측 자동화 기능을 더해 프로젝트의 목표였던 슬롯 운영 효율을 달성할 수 있었습니다.

## 특징 추출

쿠팡에 축적된 대량의 물류 데이터들로부터 트럭 수량 예측 모델에 적합한 특징들을 찾아내기 위해 탐색적 데이터 분석(EDA: Exploratory Data Analysis) 과정을 거쳤습니다. 하지만 데이터 사이에 숨겨져 있는 의미를 찾아내기 위해서는 도메인 전문가들의 조언이 필요했습니다. 물류 담당자들과 심층 인터뷰를 진행하면서 물류 입고 과정에서 나타나는 의미있는 패턴들을 다양하게 파악할 수 있었으며, 이러한 과정을 통해 트럭 수량 예측에 유용하게 활용할 수 있는 다수의 특징들을 발견하였습니다. 그리고 이렇게 도출된 특징들을 피처 엔지니어링(Feature Engineering)으로 가공해 최종적인 피처 셋(Feature Set)을 확정했습니다.

## 모델 학습: LightGBM 알고리즘

약 2 년여의 기간 동안 수집된 입고 신청 데이터로부터 약 80 만 건의 학습 데이터를 추출했습니다. 데이터 세트의 크기가 작지 않다 보니 빠른 학습과 튜닝이 가능한

알고리즘을 찾게 되었습니다. 또한 확정된 특징들 중 많은 것들이 범주형(categorical) 특징이었습니다. 이러한 특성을 갖는 데이터 세트에서 높은 예측 성능을 보이는 LightGBM 알고리즘을 사용하기로 결정했습니다.

LightGBM은 트리 기반 부스팅 모델이며, 많은 기계 학습 문제에서 성능을 입증한 알고리즘입니다. 다른 트리 기반 알고리즘들은 트리를 수평적으로 확장하는 level-wise tree growth 방식을 사용하는 반면, LightGBM은 트리를 수직적으로 확장하는 leaf-wise tree growth 방식을 사용해 학습 속도가 빠르다는 것이 특징입니다. Level-wise tree growth 방식은 각 트리 깊이가 모두 확장될 때까지 기다려야 하지만, leaf-wise 방식은 이를 기다리지 않고 가장 손실이 높은 리프 노드를 분할해 나가면서 수직으로 확장합니다. 이러한 방식 덕분에 LightGBM은 다른 알고리즘들보다 빠른 속도로 학습이 수행할 수 있습니다.

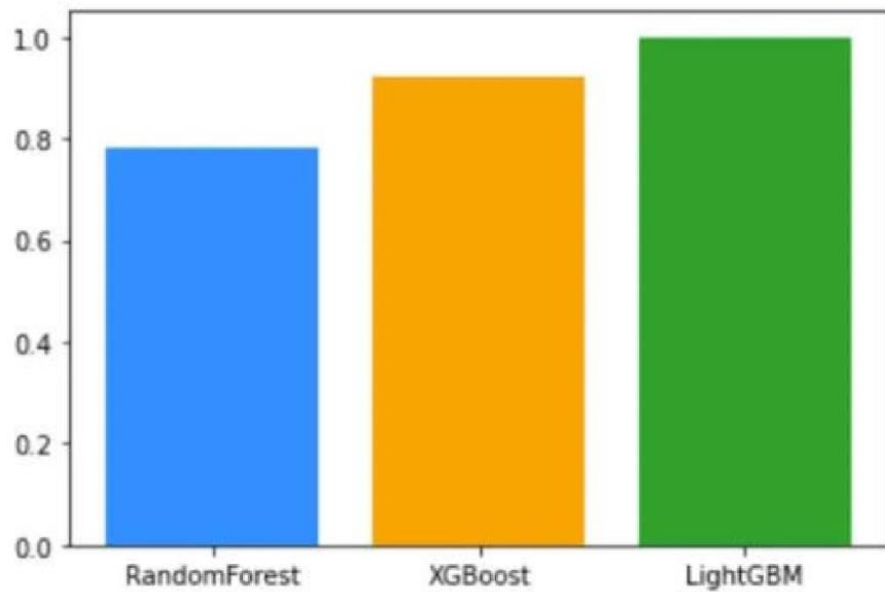
또한 다른 대부분의 알고리즘들과 달리 LightGBM은 범주형 특징에 대해 별도의 원핫 인코딩(one-hot encoding)을 해주지 않아도 됩니다.

LightGBM이 [Fisher](#) 알고리즘을 적용해 데이터의 클래스를 최적으로 분할하는 학습을 수행해 나가기 때문입니다. 이로 인해 LightGBM은 일반적으로 보다 빠른 학습 속도와 함께 높은 예측 성능을 보여줍니다. 검증(validation) 데이터 세트로 다른 주요 트리 기반 알고리즘들과 비교 평가한 결과, LightGBM은 저희가 원하는 학습 속도와 예측 성능을 보여주었습니다.

그림 2. 주요 트리(tree) 기반 알고리즘들의 검증(validation) 데이터에 대한 예측 성능 확인 결과

모델 하이퍼 파라미터 탐색 : 베이지안 최적화

모델의 하이퍼 파라미터 탐색은 [베이지안 최적화](#)(Bayesian Optimization)를 활용해 자동으로 진행되도록 구성하였습니다. 베이지안 최적화는 목적 함수  $f$ 의 함수



값  $f(x)$ 를 최대로 만드는 입력 값  $x$ 의 전역 최적해(global optimization)  $x^*$ 를 찾기 위한 방법이며, 목적 함수가 명시적이지 않고 함수 값  $f(x)$ 의 계산 비용이 클 때 최적해를 효과적으로 찾아낼 수 있습니다. 머신 러닝에서는 최적의 하이퍼 파라미터 조합을 효율적으로 찾아내는 방법 중 하나로 많이 사용됩니다. 베이지안 최적화가 이루어지는 과정을 개략적으로 설명하면 다음과 같습니다.

- (1) 미리 지정된 하이퍼 파라미터의 탐색 범위 내에서  $n$  개의 값들을 랜덤하게 선택하여 모델을 학습하고, 학습된 모델의 함수 값을 계산
- (2) 입력 값과 함수 값 쌍으로 집합을 구성하고, 이를 바탕으로 Gaussian Process 등의 방법을 사용하여 미지의 목적 함수  $f$ 를 확률적인 방법으로 추정
- (3) 현재까지 목적 함수에 대해 추정한 결과를 바탕으로,  $x^*$ 를 찾을 수 있을 것으로 예상되는 다음 입력 값 후보를 선택하여 모델을 학습하고, 학습된 모델의 함수 값을 계산하여 입력 값과 함수 값 쌍의 집합에 추가

(4) 위 (2)~(3)의 과정을 정해진 횟수만큼 반복하면서 추정 함수를 갱신한 후, 추정 함수의 함수 값을 최대로 만드는 최적 하이퍼 파라미터  $x^*$ 를 선택

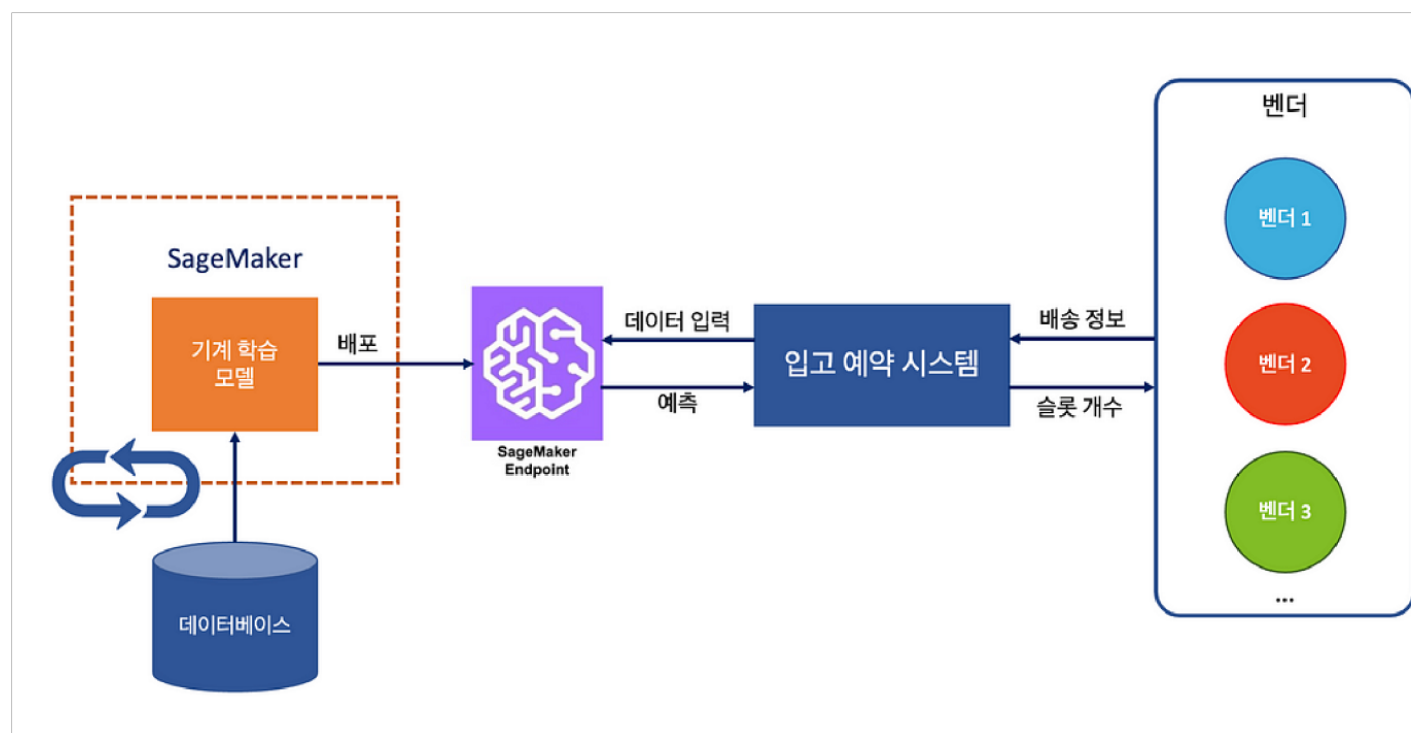
이 과정은 한 달 주기로 새로운 데이터들을 추가 반영하여 반복되면서 모델을 업데이트해 나가게 됩니다.

### 입고 예약 시스템과 모델 연계 구성

모델은 [SageMaker](#)에서 학습됩니다. 업체가 입고 예약 시스템으로 슬롯 예약 신청을 하면, 예약 시스템은 SageMaker 엔드포인트를 호출해 모델 예측 결과 값을 반환 받아 업체에게 적정 트럭 대수를 알려줍니다.

그림 3. 입고 예약 시스템과 모델 연계 구성

### 과소 예측과 과대 예측 사이의 트레이드 오프



기계 학습 예측 모델은 필연적으로 오차를 동반합니다. 저희가 해결하고자 하는 문제의 경우 실제 적정 슬롯 개수보다 적은 개수를 예측하면 과소 예측이, 많은 개수를 예측하면 과대 예측이 발생합니다. 과소 예측과 과대 예측 사이에는 트레이드 오프가



있는데, 같은 수준의 오차율 내에서 예측 모델이 적정 트럭 개수를 가급적 적게 예측하는 경향을 띠면 과대 예측은 줄어들지만 과소 예측이 늘어나고, 트럭 개수를 가급적 많이 예측하는 경향을 갖게 되면 과소 예측은 줄어들지만 과대 예측이 늘어납니다.

저희의 기본적인 목표는 과대 예측을 줄여 불필요하게 예약되는 슬롯들을 최소화하는 것이지만, 과대 예측을 지나치게 최소화하는 방향으로 모델의 예측 슬롯 개수를 결정하면, 업체들 입장에서는 너무 적은 슬롯들이 배정된다고 느끼게 되는 의도치 않은 상황이 발생할 수도 있습니다.

따라서 저희는 관련 부서들과 논의 끝에 적절한 수준의 과소 예측 비율을 유지하면서 과도 예측을 최대한 줄이는 것이 쿠팡과 업체 모두에게 이로운 방식임을 확인하였으며, 이를 반영해 최종 모델을 도출하였습니다. 모델 적용 결과

최종 학습된 모델의 적정 트럭 대수를 과소 예측하는 비율은 2.53% 수준이며, 과대 예측하는 비율은 5.04% 수준입니다. 이는 이전에 업체들이 직접 적정 트럭 대수를 예측해 입고 예약 시스템에 등록할 때의 과소 예측 비율 8.71%와 과대 예측 비율 44.45%로부터 크게 개선된 수치입니다.

이 학습 모델을 활용한 결과, 슬롯이 부족해 업체가 입고일을 변경하는 사례가 67.9% 감소했고, 결과적으로 업체는 원하는 일정으로 제품을 풀필먼트 센터로 입고할 수 있게 되었고 쿠팡은 불필요하게 낭비되는 비용을 줄이고 필요한 수량의 물품들을 원하는 일정에 납품 받을 수 있게 되었습니다.

| 구분    | 모델 적용 전 | 모델 적용 후 |
|-------|---------|---------|
| 과소 예측 | 8.71%   | 2.53%   |
| 과대 예측 | 44.45%  | 5.04%   |

표 1. 트럭 수량 예측 모델 적용 결과

향후 계획

쿠팡의 비즈니스 영역이 점점 더 확대될수록, 쿠팡이 취급하는 제품군도 다양해지고 있습니다. 최근에는 가전 제품을 전문적으로 취급하는 풀필먼트 센터들도 생겨나는 등, 쿠팡의 작업 방식도 변해가면서 이에 맞는 정확한 적정 트럭 대수 예측이 필요해진 상황입니다. 따라서 이런 새로운 유형의 제품과 작업 방식의 변화들에 대해서도 모델이 적절한 예측 성능을 보일 수 있도록, 저희는 앞으로도 파생 특징들을 발굴하고 데이터를 추가하면서 지속적으로 모델을 개선해 나갈 예정입니다.