

# Ch7. 머신러닝 면접 문제

35개 실전 면접 문제 — 출제 기업별 분석 & 풀이 전략

# 문제 구성 & 패턴 개요

35문제 (Easy 11 / Medium 14 / Hard 10)

8개 핵심 패턴으로 분류하여 학습

# 35개 면접 문제 — 패턴별 분류

패턴	대표 문제	출제 기업
회귀 & 정규화	Q7.2, 7.5, 7.9, 7.20, 7.31	Square, AQR, Two Sigma, Uber
Bias-Variance & 모델 선택	Q7.4, 7.8, 7.14, 7.15, 7.25	Salesforce, FB, Uber, Citadel
트리 & 앙상블	Q7.6, 7.10, 7.19	Point72, PWC, Microsoft
분류 & 불균형 데이터	Q7.1, 7.7, 7.12, 7.22	Robinhood, PayPal, Affirm
신경망 & 딥러닝	Q7.18, 7.21, 7.33	Amazon, Uber, Citadel
E2E ML 설계	Q7.11, 7.16, 7.24, 7.30, 7.34	DoorDash, Salesforce, Spotify
피처 엔지니어링 & 결측치	Q7.3, 7.13, 7.17	Facebook, Google, Spotify
고급: PCA, GMM, SVM	Q7.26-7.29, 7.32, 7.35	Walmart, Two Sigma, Stripe

# Pattern 1

## 회귀 & 정규화

---

Q7.2, 7.5, 7.9, 7.20, 7.31

# Q7.2 [Square] MSE vs MAE 차이점

---

## MSE (평균제곱오차)

---

- 오차를 제곱 후 평균 — 큰 오차에 높은 가중치
- 이상치(outlier)에 민감
- 그래디언트 계산 용이 → 학습 효율적
- 가우시안 노이즈 가정 시 최적 (조건부 평균 최소화)

## MAE (평균절대오차)

---

- 오차의 절대값 평균 — 이상치에 강건
- 선형 프로그래밍으로 그래디언트 계산 필요
- 조건부 중앙값(median) 최소화
- 이상치가 많거나 강건성 필요 시 선택

# Q7.5 [AQR] 다중공선성 문제와 대응

---

문제: 예측변수 간 상관관계 존재 시

- 계수 추정값·부호가 불안정하게 변동
- 중요 변수도 p-value가 높아 유의하지 않은 것처럼 보임
- 변수의 효과가 상관 변수 간 "분산"됨

해결 전략:

- 상관 변수 중 하나를 제거 (원인 분석 후)
- 교호작용항(interaction term) 추가로 결합
- 데이터 센터링 + 더 큰 표본 확보
- Ridge 회귀 등 정규화 적용

# Q7.9 [Two Sigma] 데이터 중복 시 회귀 계수 변화

---

상황: 선형 회귀 데이터를 실수로 2배 복제

풀이: 최소제곱 추정량  $\hat{\beta} = (X^T X)^{-1} X^T y$

- 데이터 2배 복제 시  $X, y$ 를 각각 2번 쌓음
- $(2X^T X)^{-1} \cdot 2X^T y$ 로 변환

$$\hat{\beta} = (2X^T X)^{-1} \cdot 2X^T y = \frac{1}{2}(X^T X)^{-1} \cdot 2X^T y = (X^T X)^{-1} X^T y$$

결론: 회귀 계수  $\hat{\beta}$ 는 변하지 않음

- 단, 표준오차는 줄어들어 p-value가 인위적으로 낮아짐

# Q7.20 [Uber] L1 vs L2 정규화

---

## L1 (Lasso)

---

- 손실함수:  $L + \lambda \sum |w_j|$
- 계수의 **절대값** 합을 페널티
- 가중치를 정확히 **0**으로 만들
- 결과: **스파스 모델** (피처 선택 효과)

## L2 (Ridge)

---

- 손실함수:  $L + \lambda \sum w_j^2$
- 계수의 **제곱** 합을 페널티
- 가중치를 0에 **가깝게** 축소 (완전 0은 드물)
- 결과: 모든 피처 유지, **분산 감소**

# Q7.31 [Two Sigma] MLE = 최소제곱법 동치 증명

---

가정:  $Y \sim \mathcal{N}(X\beta, \sigma^2 I)$  (오차항 정규분포)

우도함수의 로그:

$$\log L(\beta, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (X\beta - Y)^T (X\beta - Y)$$

$\beta$ 에 대해 미분 시:

- 첫째 항은 상수  $\rightarrow$  무시 가능
- 최대화 문제:  $\arg \max_{\beta} -\frac{1}{2\sigma^2} (X\beta - Y)^T (X\beta - Y)$
- 부호 반전:  $\arg \min_{\beta} (X\beta - Y)^T (X\beta - Y)$

결론: 정규분포 가정 하에 MLE는 잔차제곱합 최소화와 동일

# Pattern 2

## Bias-Variance & 모델 선택

---

Q7.4, 7.8, 7.14, 7.15, 7.25

# Q7.14 [Facebook] Bias-Variance 트레이드오프

핵심 공식:

$$\text{Total Error} = \text{Bias}^2 + \text{Variance} + \text{Irreducible Error}$$

구분	Bias (편향)	Variance (분산)
원인	모델이 너무 단순 (과소적합)	모델이 너무 복잡 (과적합)
예시	비선형 관계에 선형 회귀	단순 관계에 복잡한 NN
증상	학습·검증 오차 모두 높음	학습 오차 낮고 검증 오차 높음
해결	복잡한 모델 / 피처 추가	정규화 / 데이터 추가 / 단순화

- Irreducible Error: 데이터 노이즈 자체 — 모델로 줄일 수 없음
- 목표: Bias와 Variance 둘 다 낮추는 균형점 탐색

# Q7.4 [Salesforce] 이상치에 강건한 모델 만들기

---

## 이상치 원인 파악 (먼저!)

---

- 데이터 오류? 진짜 극단값? 패턴?
- 제거 전 반드시 원인 조사 필요

## 6가지 대응 전략

---

- 정규화 추가: L1/L2로 분산 감소
- 강건한 모델 선택: 트리 기반 모델은 이상치에 덜 민감
- Winsorize: 상하위 5%를 경계값으로 캡핑
- 로그 변환: 우편향 분포 시 효과적
- 손실함수 변경: MSE  $\rightarrow$  MAE 또는 Huber Loss

# Q7.15 [Uber] 교차검증 (Cross-Validation)

---

목적: 모델 성능의 진짜 수준을 측정 + 과적합 방지

K-Fold CV 절차:

1. 데이터를  $k$ 개 동일 크기 블록(fold)으로 무작위 분할
2. 각 fold  $i$ : 나머지  $k - 1$ 개로 학습, fold  $i$ 로 검증
3.  $k$ 개 검증 오차를 평균하여 최종 성능 추정

장점:

- 학습/검증 데이터 겹침 방지 → 과적합 억제
- 전용 검증셋 불필요 → 적은 데이터에서도 활용 가능

단점: 계산량 약  $k$ 배 증가

# Q7.25 [Citadel] Naive Bayes vs 로지스틱 회귀

---

## Gaussian Naive Bayes (GNB)

---

- 생성 모델:  $P(Y)$ 와  $P(X|Y)$  직접 추정
- 피처 독립 가정 (위반 시 성능 저하)
- 적은 데이터로도 학습 가능
- 최적화 불필요 (빈도 기반 추정)
- 소규모 데이터 + 강한 사전정보 시 선택

## 로지스틱 회귀

---

- 판별 모델:  $P(Y|X)$  직접 학습
- 피처 간 상관 허용, 교호작용 약함
- 최적화 필요 (가중치 학습)
- 해석 용이 (계수  $\rightarrow$  확률)
- 데이터 충분 + 피처 상관 존재 시 선택

# Pattern 3

## 트리 & 앙상블 방법

---

Q7.6, 7.10, 7.19

# Q7.6 [Point72] 랜덤 포레스트의 핵심 아이디어

---

## 단일 결정 트리의 문제

---

- 과적합 경향이 강함
- 학습 데이터에 민감 → 불안정

## 랜덤 포레스트의 2가지 개선

---

- 배깅(Bootstrap Aggregation): 리샘플링된 데이터로 여러 트리 학습 → 분산 감소
- 피처 랜덤 선택: 각 분할에서  $m < p$  개 피처만 사용 → 트리 간 비상관화
  - 중요 피처가 항상 첫 분할에 오는 것을 방지

추가 장점:

# Q7.10 [PWC] Gradient Boosting vs Random Forest

## 공통점

- 결정 트리의 앙상블
- 전처리 부담 적음
- 유연한 모델

## 차이점

항목	Gradient Boosting	Random Forest
학습	순차적 (약한 학습기가 이전 오류 학습)	병렬적 (독립 트리)
결합	반복마다 누적	최종에 평균/투표

항목

과제

튜닝

적합

# Q7.19 [Microsoft] 엔트로피와 정보이득

엔트로피: 샘플의 불순도 측정

$$\text{Entropy} = - \sum_{i=1}^k P(Y = l_i) \log_2 P(Y = l_i)$$

- 완전 동질 → Entropy = 0 / 50-50 분할 → Entropy = 1

정보이득: 속성 분할 후 엔트로피 감소량

$$IG(S, Y) = H(Y) - H(Y|S)$$

수치 예시: a 5개 + b 5개 → 속성 X로 분할

- $X = 1$ : a 5개, b 1개 /  $X = 0$ : b 4개
- $H(\text{After}) = \frac{4}{10} (0) + \frac{6}{10} (0.65) = 0.39$
- $IG = 1 - 0.39 = 0.61 \rightarrow$  좋은 분할

# Pattern 4

## 분류 & 불균형 데이터

---

Q7.1, 7.7, 7.12, 7.22

# Q7.1 [Robinhood] 불균형 데이터셋 (1%:99%) 처리

---

- ① 추가 데이터 확보가능성 먼저 검토
- ② 평가 지표 변경: Accuracy 대신 Precision, Recall, F1, ROC 사용
- ③ 리샘플링: 소수 클래스 오버샘플링 or 다수 클래스 언더샘플링 (부트스트래핑)
- ④ 합성 데이터 생성: SMOTE — 이웃 인스턴스 간 속성을 보간하여 합성 샘플 생성
- ⑤ 앙상블 활용: 클래스 비율 달리한 여러 모델 또는 소수 클래스 전체 + 다수 일부 조합
- ⑥ 분류 임계값 조정: 0.5 대신 비율에 맞는 threshold 설정
- ⑦ 커스텀 손실함수: 소수 클래스 오분류에 더 큰 페널티 부여

# Q7.7 [PayPal] 결측값이 많은 사기 탐지 데이터

---

## Step 1 — 결측 유형 파악:

- MCAR / MAR / NMAR 여부 → 보간 방법 결정
- 어떤 피처에 결측? 수치형 vs 범주형? 패턴 존재?

## Step 2 — 베이스라인 우선 구축:

- 결측 데이터가 예측에 중요한지 먼저 확인
- 결측 없는 컬럼만으로 비즈니스 목표 달성 가능 여부 점검

## Step 3 — 보간(Imputation):

- 연속형: 평균/중앙값 → 더 정교하게는 KNN 기반 보간
- 교차검증으로 보간 전후 성능 비교

## Step 4 — 외부 데이터: 제3자 데이터셋으로 결측 보완 가능성 탐색

# Q7.12 [Affirm] 대출 거절 사유 설명 (가중치 없이)

---

핵심 기법: Partial Dependence Plot (PDP)

- 모델 학습 후, 각 피처를 개별적으로 변화시키며 예측값 관찰
- 다른 피처는 고정 → 해당 피처의 한계 효과(marginal effect) 파악

구체적 예시:

- 4명의 신청자: 소득·부채·카드 수 동일, FICO 점수만 상이
- FICO 700, 720 → 승인 / FICO 600, 650 → 거절
- 나머지 변수 통제됨 → "FICO 점수가 낮아서 거절" 이라고 설명 가능

장점: 모델 내부 가중치를 들여다보지 않아도 근거 기반 설명 가능

## Q7.22 [Affirm] 점수에 제곱근 적용 시 ROC 변화

---

상황: 사기 확률 점수  $s \in [0, 1]$ 에  $\sqrt{s}$  적용

ROC 곡선은 변하지 않는다!

- ROC는 \*\*상대적 순서(ranking)\*\*만 중요
- $Y > X$ 이면  $\sqrt{Y} > \sqrt{X}$ 도 성립 (단조증가)
- 임계값만 바뀌고, TPR/FPR 조합은 동일

ROC가 변하는 함수: 단조증가가 아닌 함수

$$f(x) = -x, \quad f(x) = -x^2, \quad \text{계단함수}$$

- 이들은 상대적 순서를 뒤집거나 뭉개므로 ROC 변동

# Pattern 5

## 신경망 & 최적화

---

Q7.18, 7.21, 7.23, 7.33

# Q7.18 [Amazon] 볼록 함수와 신경망의 비볼록성

---

볼록 함수 정의: 임의의 두 점  $x, y$ 에 대해

$$f((1 - \alpha)x + \alpha y) \leq (1 - \alpha)f(x) + \alpha f(y), \quad 0 < \alpha < 1$$

- 두 점을 잇는 선분이 항상 함수 그래프 위에 위치
- 모든 극소(local min) = 전역 최소(global min) → 최적화 보장

신경망은 왜 비볼록(Non-convex)인가?

- NN은 범용 함수 근사기 → 비볼록 함수도 근사 가능
- 비볼록 함수를 볼록 함수로 근사할 수 없음 → NN 자체가 비볼록
- 노드를 서로 교환해도 동일 출력 → 다수의 local minima 존재
- 특정 global minimum이 없으므로 볼록 불가

# Q7.21 [Amazon] 경사하강법과 SGD

---

경사하강법(GD): 최적화할 함수의 급경사 방향으로 이동

$$x_{t+1} = x_t - \alpha \nabla f(x_t)$$

- $\alpha$ : 학습률,  $\nabla f$ : 함수의 그래디언트
- 수렴까지 반복

확률적 경사하강법(SGD):

- 전체 데이터가 아닌 무작위 1개 샘플로 그래디언트 추정
- $E[g(x)] = \nabla f(x) \rightarrow$  비편향 추정량
- 대규모 데이터에서 훨씬 빠름
- 국소 최소/안장점 탈출에 유리

GD vs SGD: 정확도 vs 속도 트레이드오프 — 실무에서는 미니배치 SGD 주류

## Q7.23 [IBM] 가우시안 확률변수의 엔트로피

---

설정:  $X \sim \mathcal{N}(\mu, \sigma^2)$

연속 확률변수 엔트로피:

$$h(X) = - \int p(x) \log p(x) dx$$

가우시안 대입 후 전개:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$
$$h(X) = - \int p(x) \log \frac{1}{\sigma\sqrt{2\pi}} dx - \int p(x) \cdot \left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \log e dx$$

결과:  $H(X) = \frac{1}{2} + \log(\sigma\sqrt{2\pi})$

- 분산  $\sigma^2$ 이 클수록 엔트로피 증가 (불확실성 큼)

# Q7.33 [Citadel] 로지스틱 회귀의 수리적 정식화

모델:  $k$ 개 클래스에 대해

$$\log \frac{P(C = j | X = x)}{P(C = K | X = x)} = \beta_{j0} + \beta_j^T x$$

이진 분류 로그우도 ( $p(x) = P(C = 1 | X, \beta)$ ):

$$L(\beta) = \sum_{i=1}^N [y_i \log p(x_i) + (1 - y_i) \log(1 - p(x_i))]$$

최적화:

- $\frac{\partial L}{\partial \beta}$  를 구하고 0으로 설정
- 닫힌 해(closed-form) 없음 → 반복 최적화 (Newton-Raphson 등)
- 볼록 함수이므로 전역 최적 수렴 보장

# Pattern 6

## End-to-End ML 시스템 설계

---

Q7.11, 7.16, 7.24, 7.30, 7.34

# Q7.11 [DoorDash] 10K 데이터로 ETA 모델 충분 한가?

## Step 1: 명확화

- "정확한" 기준이 뭐가?
- ETA 용도: 고객 표시 vs 배차 알고리즘?
- 과소 추정 vs 과대 추정 비즈니스 영향

## Step 2: 베이스라인

- 단순 모델: 운전시간 + 평균 준비시간
- RMSE, MAE,  $R^2$ 로 평가

## Step 3: 학습곡선

- 데이터 25%, 50%, 75%로 성능 비교
- 감소폭 둔화 → 피처 개선이 더 효과적

## Step 4: 데이터 부족 시

- 피처 추가 (수요·공급, 교통)
- 차원 축소 / 피처 선택
- 유사 시장 데이터 활용

# Q7.16 [Salesforce] 리드 스코어링 알고리즘 설계

## Step 1: 요구사항

- 자사용 vs 제품 내장?
- 설명 가능성 필요?
- CRM 내부만 or 외부 전망?

## Step 2: 피처

- Firmographic (산업, 매출, 직원수)
- 마케팅 활동 (클릭, 다운로드)
- 세일즈 미팅 빈도/최근성
- 딜 상세 (제품, 규모, 계약)

## Step 3: 모델

- 로지스틱: 해석 쉬움, 복잡 교호작용 약함
- NN/SVM: 고차원, 설명 어려움
- 트리 기반 (RF/XGBoost): 최적 절충

## Step 4: 배포

- 피처 드리프트 모니터링
- 고객 프로필 변화에 따른 재학습

# Q7.30 [Robinhood] 이탈(Churn) 예측 모델

## Step 1: 이탈 정의

- Gold 구독 취소?
- 거래 활동 장기 중단?
- 잔고 \$10 이하 유지?
- 2% 월 이탈 = 27% 연 이탈

## Step 2: 모델 선택

- 확률 출력: 로지스틱 회귀
- 설명 필요: 결정 트리 / RF
- 탐지만: NN / SVM도 가능

## Step 3: 피쳐

- 잔고 (절대값 + 추세)
- 대규모 손실 경험
- DAU → 사용 빈도 감소
- 인구통계 정보

## Step 4: 배포

- 오프라인 지표 검증 후
- A/B 테스트로 영향 측정
- 오류 분석 + 지속 재학습

# Q7.24 [Stitch Fix] 구매 성향(Propensity) 모델

## 접근법

- 변수: 구매 여부(종속) + 연령/성별/소득 등 (독립)
- 이진 분류 → **확률 예측 모델**

## 모델 비교

모델	장점	단점
로지스틱	해석 용이, 확률 직접 출력	교호작용 못 잡음, 불안정
NN/SVM	고차원, 복잡 패턴	설명 어려움, 대량 데이터 필요
RF/트리	정확도 + 해석력 균형	극단적 고차원 약함

# Q7.34 [Spotify] Discover Weekly 추천 알고리즘

---

Step 1 — 명확화: 새 음악 발견 vs 좋아할 곡? (탐색 vs 활용 트레이드오프)

Step 2 — 피쳐 설계:

- 핵심: 사용자-곡 인터랙션 (재생 횟수를 암묵적 평점으로)
- 곡 메타데이터: 아티스트, 앨범, 플레이리스트 소속
- 오디오 피쳐: 템포, 악기, 보컬 비율
- 인구통계: 연령, 지역 (지역적 취향 차이)

Step 3 — 협업 필터링:

- 사용자-곡 행렬  $R = PQ^T$  (행렬 분해)
- ALS(교대 최소제곱) → 분산 컴퓨팅으로 스케일링

# Pattern 7

## 피처 엔지니어링 & NLP

---

Q7.3, 7.8, 7.13, 7.17

# Q7.3 [Facebook] K-Means에서 k 선택법

---

## 방법 1: Elbow Method

- $x$ 축: 클러스터 수  $k$ ,  $y$ 축: 군집 내 제곱합(WCSS)
- $k$  증가 시 WCSS 급감 → 어느 지점에서 완만 → "팔꿈치" 지점이 최적  $k$
- 주의: 명확한 elbow가 없을 수도 있음

## 방법 2: Silhouette Method

- 포인트가 자기 군집 vs 가장 가까운 이웃 군집에 얼마나 잘 속하는지 측정

$$s = \frac{b - a}{\max(a, b)}$$

- $s \in [-1, 1]$ : 1이면 완벽한 군집, -1이면 잘못된 군집
- WCSS보다 계산 비용 높음

방법 3: 비즈니스 지관 — 드메인 저무가이 기대 그르 스 환용

# Q7.8 [Airbnb] 로지스틱 회귀 성능 개선 전략

---

## 모델 내 개선:

- 피쳐 정규화: 특정 변수의 과도한 가중치 방지
- 피쳐 추가: 로지스틱 회귀는 high bias → 더 많은 피쳐가 도움
- 이상치 처리: 식별 후 유지/제거 결정
- 변수 선택: 노이즈 유발 피쳐 제거
- 교차검증 + 하이퍼파라미터 튜닝: 정규화 페널티 등

## 모델 교체 고려:

- 클래스가 선형 분리 불가능할 수 있음 (로지스틱의 근본 한계)
- 대안: SVM (커널 트릭), 트리 기반 (비선형 경계), 신경망 (복잡 패턴)

# Q7.13 [Google] 대규모 코퍼스에서 동의어 식별

---

Step 1 — 단어 임베딩 생성:

- Word2Vec: 단어의 문맥 기반 벡터 생성
- 의미적으로 유사한 단어 → 벡터 공간에서 가까이 위치

Step 2 — 유사도 측정:

- 코사인 유사도 또는 유클리드 거리로 벡터 간 거리 계산

Step 3 — 동의어 추출:

- K-means 클러스터링: 임베딩 공간에서 단어 그룹핑
- KNN: 특정 단어의 최근접 이웃 = 동의어 후보

주의: 반의어 문제

- "뜨거운"과 "차가운"은 유사한 문맥에 등장 → 벡터가 가까움

# Pattern 8

**고급: PCA, GMM, SVM, MLE**

---

Q7.26-7.29, 7.32, 7.35

# Q7.26 [Walmart] K-Means 손실함수와 업데이트 규칙

손실함수:  $k$ 개 군집,  $n$ 개 샘플에 대해

$$J = \sum_{j=1}^k \sum_{x \in S_j} \|x - \mu_j\|^2$$

미분 후 업데이트 규칙:

방법	업데이트
Batch GD	$\mu_k = \mu_k + \alpha \cdot \frac{1}{ S_k } \sum_{x_i \in S_k} 2(x_i - \mu_k)$
SGD	$\mu_k = \mu_k + \alpha(x_i - \mu_k)$

- Batch GD: 모든 소속 포인트의 평균 방향으로 이동

# Q7.27 [Two Sigma] SVM 커널 트릭

---

## 핵심 아이디어

---

- 현재 차원에서 선형 분리 불가능한 데이터를
- 고차원 공간으로 매핑하면 분리 가능
- $k(x, y) = \phi(x)^T \phi(y)$

## 2차 커널 예시

---

- $k(x, y) = (1 + x^T y)^2 = \phi(x)^T \phi(y)$
- 원래 공간에서 계산  $\rightarrow \phi$  명시적 변환 불필요
- 계산 폭발 없이 고차원 매핑 효과

## 커널 선택 가이드

---

- 선형 문제: Linear / Logistic 커널

# Q7.28 [Morgan Stanley] 가우시안 MLE 파라미터 추정

---

설정:  $x_1, \dots, x_n \sim \mathcal{N}(\mu, \sigma^2)$  (i.i.d.)

로그 우도:

$$\log p(X|\mu, \sigma^2) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 - \frac{n}{2} \log \sigma^2 - \frac{n}{2} \log \pi$$

$\mu$ 에 대해 미분  $\rightarrow 0$ :

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i \quad (\text{표본평균})$$

$\sigma^2$ 에 대해 미분  $\rightarrow 0$ :

$$1 - \frac{n}{2\sigma^2}$$

# Q7.29 [Stripe] GMM 기반 이상 거래 탐지

모델 설정:  $K$  개 클래스의 가우시안 혼합

$$p(x) = \sum_{k=1}^K \pi_k \mathcal{N}(x | \mu_k, \Sigma_k), \quad \sum_k \pi_k = 1$$

사후 확률 (Bayes' Rule):

$$P(k|x) = \frac{\pi_k \mathcal{N}(x | \mu_k, \Sigma_k)}{\sum_j \pi_j \mathcal{N}(x | \mu_j, \Sigma_j)}$$

학습: EM 알고리즘으로  $\{\mu_k, \Sigma_k, \pi_k\}$  반복 추정

이상 탐지 로직:

- 새 거래  $x_{new}$  에 대해 모든  $K$  개 클래스의 사후 확률 계산
- 모든 클래스에서 확률이 낮으면 → 어떤 정상 패턴에도 불일치 → 사기 의심

# Q7.32 [Uber] PCA 정식화와 유도

목표:  $p$ 차원 데이터  $x$ 를  $k \ll p$  차원으로 축소 (분산 최대화)

선형 결합:  $y_i = w_i^T x$ , 제약  $w_i^T w_i = 1$  (정규직교)

라그랑주 승수법:

$$\frac{\partial}{\partial w} [w^T \Sigma w - \lambda(w^T w - 1)] = 0$$

$$\Sigma w = \lambda w$$

- $w$ 는 공분산 행렬  $\Sigma$ 의 고유벡터(eigenvector)
- $\lambda$ 는 대응하는 고유값(eigenvalue)

결론:

- $\text{Var}(y) = w^T \Sigma w = \lambda \rightarrow$  분산 = 고유값

가장 큰 고유값의 고유벡터가 제1 주성분

# Q7.35 [Google] 최소제곱 추정량의 분산-공분산 행렬

출발:  $\hat{\beta} = (X^T X)^{-1} X^T y$ , 가정  $y = X\beta + \varepsilon, \varepsilon \sim \mathcal{N}(0, \sigma^2)$

대입:

$$\hat{\beta} = (X^T X)^{-1} X^T (X\beta + \varepsilon) = \beta + (X^T X)^{-1} X^T \varepsilon$$

분산 계산:  $\text{Var}(\hat{\beta}) = E[\hat{\beta}^2] - E[\hat{\beta}]^2$

- $E[\hat{\beta}] = \beta$  (비편향)이므로:

$$\begin{aligned} \text{Var}(\hat{\beta}) &= E[((X^T X)^{-1} X^T \varepsilon)^2] \\ &= (X^T X)^{-1} X^T X (X^T X)^{-1} \cdot E[\varepsilon^2] \end{aligned}$$

$$\boxed{\text{Var}(\hat{\beta}) = (X^T X)^{-1} \sigma^2}$$

# 전체 요약

## 8개 패턴 핵심 정리

---

# ML 면접 8대 패턴 — 핵심 키워드

## 1. 회귀 & 정규화

MSE vs MAE, 다중공선성, L1(스파스) vs L2(축소),  
MLE=OLS 동치

## 2. Bias-Variance

편향-분산 공식, 이상치 6전략, 교차검증 K-fold

## 3. 트리 & 앙상블

RF = 배깅+피쳐 랜덤, GB vs RF 차이, 정보이득/엔트로피

## 4. 분류 & 불균형

SMOTE, 결측치 MCAR/MAR/NMAR, PDP로 설명,  
ROC 불변성

PCA

## 5. 신경망 & 최적화

볼록/비볼록, GD vs SGD, 로지스틱 회귀 MLE

## 6. E2E 설계

명확화 → 베이스라인 → 학습곡선 → 배포 모니터링

## 7. 피쳐 & NLP

Elbow/Silhouette, Word2Vec 동의어, 협업 필터링

## 8. 고급 수리

# 면접 문제 풀이 프레임워크

모든 ML 면접 문제에 적용 가능한 4단계:

단계	핵심 질문	출제 빈도 높은 포인트
Clarify	정확히 뭘 예측? 지표는? 사용 맥락?	DoorDash ETA, Robinhood Churn
Baseline	가장 단순한 모델로 먼저 성능 확인	10K 데이터 충분성 검증
Model	해석 가능성 vs 정확도 트레이드오프	로지스틱 → 트리 → NN 스펙트럼
Deploy	모니터링, A/B 테스트, 재학습 계획	피처 드리프트, 콜드 스타트

핵심: "정답 모델"보다 체계적 사고 과정을 보여주는 것이 합격 포인트

# 난이도별 출제 경향 & 전략

난이도	특징	전략
Easy	개념 정의 + 비교 (MSE vs MAE, RF 동기)	정확한 정의 + 실용적 사례 1개
Medium	개념 응용 + 설계 (리드 스코어링, 추천)	단계별 프레임워크 적용
Hard	수학적 유도 + 증명 (PCA, MLE, GMM)	핵심 수식 암기 + 직관적 해석 병행

## 빈출 기업 유형:

- 금융 (Two Sigma, Citadel, Morgan Stanley): 수리 통계 + 정규화
- 테크 (Google, Facebook, Amazon): 스케일 + 시스템 설계
- 핀테크 (Affirm, Robinhood, PayPal): 불균형 데이터 + 설명 가능성