Chapter 7: Machine Learning

목차

A. ML 인터뷰 유형과 전략

ML 인터뷰 3가지 유형

Conceptual

- ML 개념/용어의 이해도 테스트
- "Bias-variance tradeoff란?"
- "PCA는 어떻게 작동하는가?"
- ELI5 방식으로 설명 요구 빈출

- 대부분의 DS 인터뷰에서 출제됨

Resume-Driven

- 실무 경험 중심 질문
- 이력서에 적은 프로젝트를 깊이 파고듦
- "사용한 모델의 trade-off는?"
- 프로젝트 외 관련 분야까지 확장됨

인터뷰 전략: 알아야 할 핵심 포인트

- **대부분의 DS는 비즈니스 문제 해결** 목적 — 딥러닝 전문성보다 기본기가 중요함
- 모르는 기법이 나오면 솔직히 인정 — "배우고 싶다"는 자세가 거짓 답변보다 나음
- **"좋아하는 ML 알고리즘은?"** 질문 빈출 — 실제 사용 경험이 있는 기법 준비 필수
- SOTA 트랜스포머보다 **기본적이면서 흥미로운 기법**이 대화에 유리함
- ML Engineering/Research Scientist 역할은 수학적 유도 + 후속 질문까지 준비 필요

B. 모델 평가와 선택

모델 평가 vs 모델 선택

- **모델 평가(Evaluation)**: 훈련 후 테스트셋에서 모델 성능을 측정하는 과정
  - 훈련 데이터(80%)와 테스트 데이터(20%)를 분리하는 것이 핵심
  - 모델의 가치는 **이전에 보지 못한 데이터**에 대한 예측력으로 결정됨
- **모델 선택(Selection)**: 평가된 모델들 중 최적 모델을 고르는 과정
  - 비즈니스/제품 제약을 함께 고려해야 함
  - Facebook에서 CTR 0.1% 개선 = $10M+ 추가 매출
- 인터뷰 케이스 스터디에서 모델 비교·대조·선택 논의가 자연스럽게 이어짐

Bias-Variance Trade-off: 개념 분해

추정 함수 $\hat{f}(x)$로 목표 변수 y를 예측할 때:

$$y = f(x) + \epsilon$$

예측 오차는 3가지 성분으로 분해됨:

성분	의미	방향
Bias	예측값이 실제 $f(x)$에서 얼마나 벗어나는가	낮을수록 좋음
Variance	훈련 데이터에 따라 예측이 얼마나 변하는가	낮을수록 좋음
Irreducible Error	데이터 자체의 노이즈 (ϵ)	줄일 수 없음

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

Bias-Variance: 실전 직관

High Bias, Low Variance

- 예: **선형 회귀**
- 구현이 쉽고 안정적임
- 주택 가격 예측 시: 예측값이 시장가와 자주 차이남
- 하지만 예측의 **분산은 낮음**
- 해결: 모델 복잡도 증가

High Variance, Low Bias

- 예: **신경망**
- 예측값이 실제에 가까움
- 하지만 입력에 따라 예측이 **크게 변동**함
- 해결: 데이터 추가 확보
- 정규화(Regularization) 적용

면접관이 정말 테스트하는 것

- 수식 유도보다 **상황에 맞는 추론 능력**이 더 자주 평가됨
- "이 모델이 high variance라면 어떻게 하겠는가?" → **데이터 추가 확보** 언급
- "이 모델이 high bias라면?" → **모델 복잡도 증가** 논의
- 비즈니스/제품 요구사항을 이해하고 **적절한 trade-off** 결정하는 것이 핵심
- 핵심 질문: "이 상황에서 bias를 줄이는 것과 variance를 줄이는 것 중 어느 것이 더 중요한가?"

Model Complexity와 Overfitting

> "All models are wrong, but some are useful" — George Box

- **Occam's Razor**: 단순한 모델이 일반적으로 더 유용하고 정확함
- 단순한 모델 → **일반화(generalize)** 능력이 높음

상태	설명	결과
Overfitting	훈련 데이터에 너무 밀착	노이즈까지 학습 → 새 데이터에서 성능 저하
Underfitting	데이터의 진짜 관계를 충분히 학습 못함	훈련/테스트 모두 성능 부족
Good Balance	적절한 복잡도	일반화 성능 최적

- 면접에서 빈출: "Overfitting을 어떻게 감지하는가?", "어떻게 방지하는가?"

C. 정규화와 해석 가능성

Regularization: L1 vs L2

L1 (Lasso)

- 계수의 **절대값**을 패널티로 추가
- 많은 계수를 **정확히 0으로** 축소함
- Feature Selection 효과 — 변수 자동 제거
- 더 **희소(sparse)한 모델** 생성
- 엄격한 축소 연산

L2 (Ridge)

- 계수의 **제곱값**을 패널티로 추가
- 계수를 **0에 가깝게** 축소하지만 완전히 0이 되진 않음
- 모든 변수를 유지하면서 영향력을 줄임
- 다중공선성 문제에 효과적
- L1보다 덜 엄격한 축소

Elastic Net과 정규화 선택 기준

- **Elastic Net**: L1과 L2를 **선형 결합**한 정규화 방식

$$\text{Loss} = \text{RSS} + \lambda_1 \sum |\beta_j| + \lambda_2 \sum \beta_j^2$$

상황	추천 방법
불필요한 변수가 많고, 변수 선택이 필요	L1 (Lasso)
상관된 변수가 많고, 모두 유지하고 싶을 때	L2 (Ridge)
변수 선택 + 상관 변수 동시 처리 필요	Elastic Net

- Bias-Variance 관점: 정규화는 variance를 **크게 줄이면서** bias를 **약간만** 증가시킴

모델 해석 가능성 (Interpretability)

- Kaggle에서는 정확도만 최적화하지만, **현실에서는 설명 가능성도 필수**임
- 대출 거절 시 이유 설명 의무, 의료 분야 감사 대응, AI 편향 탐지 등

모델 유형	해석 방법
선형 모델	가중치(weights) 시각화 및 분석
랜덤 포레스트	Feature Importance 내장
블랙박스 모델	SHAP, LIME 등 외부 프레임워크

- **SHAP**: Shapley 값 기반 — 각 피처의 평균 한계 기여도
- **LIME**: 개별 예측 주변의 희소 선형 모델로 로컬 해석
- 면접에서 SHAP/LIME 세부사항보다 **"왜 해석 가능성이 중요한가"**가 핵심

D. 모델 훈련과 교차 검증

모델 훈련의 기본 구조

- 기본: 훈련 데이터(80%) + 테스트 데이터(20%)
- 하지만 단순 80/20 분할 이상의 기법이 필요함

개념	설명
Gradient Descent	손실 함수를 최소화하는 최적화 방법 — 가장 가파른 방향으로 이동
SGD	데이터 1개씩 사용 — 로컬 최소값 탈출 가능, 계산 효율적
Mini-batch GD	소규모 배치 사용 — GD와 SGD의 절충

경사 하강법 업데이트 규칙:

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

- α_t: 학습률(learning rate) — 스텝 크기를 결정함

K-Fold 교차 검증 절차

❶ 데이터를 k개의 동일 크기 블록(fold)으로 **무작위 분할**

❷ 각 fold i에 대해: fold i를 제외한 나머지로 훈련 → fold i로 **검증 오차** 측정

❸ k개 검증 오차를 **평균**하여 모델의 진정한 오차를 추정함

교차 검증 변형과 시계열 주의점

주요 변형

- K-Fold CV: 가장 보편적 (보통 k=5 또는 10)
- LOOCV: $k = n$ (데이터 수만큼 fold) — 소규모 데이터에 적합
- Train-Validation-Test Split: 대규모 데이터에서 계산 효율적 — 별도 validation set(10~20%) 설정

시계열 CV 주의

- 표준 K-Fold CV를 **시계열에 그대로 적용하면 안 됨**
- 시계열은 시간 순서가 있음 — "미래" 데이터로 "과거"를 예측하면 안 됨
- **해결**: 시간 기준점을 시작~끝까지 이동하며 과거 데이터만으로 훈련
- 면접 빈출 질문: "시계열에 CV를 어떻게 적용하는가?"

Bootstrapping, Bagging, 하이퍼파라미터 튜닝

- **Bootstrapping**: 대규모 샘플에서 **복원 추출**로 반복 샘플링 → 모집단 추정
  - 소규모 데이터, 클래스 불균형 해결에 유용함
- **Bagging**: Bootstrap + Aggregation — 여러 소규모 모델을 **앙상블**하여 결합
  - 랜덤 포레스트의 핵심 메커니즘 (Part 2에서 상세 다룸)

튜닝 방법	특징
Grid Search	모든 조합 시도 — 확실하지만 조합 폭발 문제
Random Search	무작위 샘플링 — 효율적이지만 최적 보장 없음
Bayesian Optimization	사전 확률 기반 탐색 — ML Engineering 역할에서 출제

Learning Curve로 Overfitting 감지

- **학습 곡선**: y축 = 성능 지표(예: 정확도), x축 = 반복 횟수(경험)

패턴	해석
훈련 오차 ↓ + 검증 오차 ↓	정상 학습 진행 중
훈련 오차 ↓ + 검증 오차 ↑	Overfitting — 훈련 중단 필요
훈련·검증 사이 큰 갭이 줄지 않음	데이터가 **비대표적**임

- 면접 빈출: "모델이 overfitting하는지 어떻게 식별하겠는가?"
- 학습 곡선 분석으로 **조기 종료(early stopping)** 시점을 판단할 수 있음

E. 선형 회귀

선형 회귀: 왜 가장 많이 물어보는가

- 빠른 실행 시간 + 높은 해석 가능성 → **실무에서 가장 많이 사용됨**
- "Regression to regression": 고급 기법 시도 후 결국 선형 회귀로 돌아오는 현상

$$\hat{y} = X\beta$$

- X: 예측 변수 행렬, β: 각 변수의 가중치 벡터
- 면접관은 `sklearn`의 `fit()` 호출 이상의 **깊은 이해**를 평가함
  - 가정(assumptions), 실전 엣지 케이스, 평가 지표 숙지가 차별화 포인트

선형 회귀 평가 지표

잔차(Residual): 예측값과 실제값의 거리

$$\mathrm{RSS}(\beta) = (y - X\beta)^T (y - X\beta)$$

지표	수식	특징
RSS	$\sum (y_i - \hat{y}_i)^2$	최소화 대상
TSS	$\sum (y_i - \bar{y})^2$	데이터 총 변동 = ESS + RSS
R^2	$1 - \frac{RSS}{TSS}$	0~1, 모델이 설명하는 변동 비율
MSE	$\frac{1}{n}\sum (y_i - \hat{y}_i)^2$	큰 오차에 민감 (이상치에 취약)
MAE	$\frac{1}{n}\sum	y_i - \hat{y}_i

R^2의 함정과 모델 복잡도 조정

- **"변수 추가 시 R^2에 어떤 영향이 있는가?"** — 대표적 면접 질문
- 변수를 추가하면 R^2은 **항상 증가** → 하지만 overfitting 위험도 증가
- R^2만으로 모델 품질을 판단하면 안 됨

보정 지표	핵심
Adjusted R^2	변수 수에 페널티 부여
AIC	모델 복잡도와 적합도의 균형
BIC	AIC보다 더 강한 복잡도 패널티
Mallow's C_p	편향 추정의 대리 지표

Subset Selection: 변수 선택 전략

- 기본적으로 모든 예측 변수를 사용하지만, 실무에서는 **핵심 변수만 선택**해야 함

방법	프로세스	장단점
Best Subset	p개 중 k개 조합을 모두 시도	최적 보장이나 p 증가 시 계산 불가능
Forward Stepwise	빈 모델 → 가장 유용한 변수를 순차 추가	효율적, 최적 미보장
Backward Stepwise	전체 모델 → 가장 불필요한 변수를 순차 제거	효율적, 최적 미보장

- 선택 기준: 높은 R^2 + 낮은 RSS + AIC/Adjusted R^2 고려

선형 회귀 4대 가정

1. **선형성(Linearity)**: 피처와 목표 변수 사이의 관계가 선형이어야 함
2. **등분산성(Homoscedasticity)**: 잔차의 분산이 일정해야 함
3. **독립성(Independence)**: 모든 관측값이 서로 독립이어야 함 (i.i.d.)
4. **정규성(Normality)**: Y의 분포가 정규분포를 따라야 함 (i.i.d.)

가정 위반 시 발생하는 문제

- 4개 가정 중 **하나라도 위반되면**: 예측과 신뢰구간이 편향되거나 오해를 유발함
- 동일한 최적적합선(line of best fit)이라도 **가정 충족 여부에 따라 의미가 달라짐**

위반 유형	감지 방법	대응
이분산성	잔차 vs 적합값 플롯 (비선형 패턴)	종속변수 변환, 비선형 항 추가
비정규성	QQ Plot (직선 이탈 여부)	로그/제곱근 변환
이상치	Cook's Distance (영향력 추정)	임계값 초과 포인트 제거
다중공선성	VIF (분산팽창계수)	변수 제거, PCA, PLS

이분산성과 QQ Plot 진단

이분산성 (Heteroscedasticity)

- 잔차의 분산이 일정하지 않은 상태
- **잔차 vs 적합값** 플롯에서 비선형 패턴이 보이면 의심
- **Scale-Location 플롯**: 표준화 잔차 vs 적합값 — 수평선이 아니면 이분산성
- 대응: 종속변수 변환 또는 비선형 항 추가

QQ Plot 해석

- 표준화 잔차 vs 이론적 분위수 그래프
- **직선**: 정규분포 충족
- **끝이 위로 휘면**: Heavy-tailed
- **끝이 아래로 휘면**: Light-tailed
- **S자 곡선**: 왜도(skew) 존재
- 직선에서 벗어나면 → 변환 필요

다중공선성과 교란 변수

- **다중공선성(Multicollinearity)**: 예측 변수들이 서로 상관되어 가중치 추정이 왜곡됨
  - 예: 인스타 게시물 수 ↔️ 알림 수 (둘 다 사용자 활동에 연관)
  - VIF로 진단 → 변수 제거, 결합, 또는 PCA 적용
- **교란 변수(Confounding)**: 독립/종속 변수 모두에 영향을 미치는 제3의 변수
  - 예: 아이스크림 소비 → 화상? 아니, **기온**이 교란 변수임
  - Selection Bias: 데이터 수집 방식의 편향
  - Omitted Variable Bias: 중요 변수 누락 → 편향된 모델
  - 대응: **층화(Stratification)** + 카이제곱 검정

일반화 선형 모델 (GLM)

- 선형 회귀의 가정(잔차 정규분포)을 **완화**한 일반화 버전

GLM 구성 요소	역할	예시
Random Component	오차항의 분포	정규분포, 포아송, 이항
Systematic Component	예측 변수의 선형 결합	$\beta_0 + \beta_1 x_1 + \dots$
Link Function	Random ↔ Systematic 연결	항등, 로짓, 로그

- 예: Tinder가 월별 매칭 수를 예측 → **포아송 회귀**(카운트 데이터)가 적합함
- 비선형 회귀(다항식, 스플라인, GAM)도 존재하나 면접에서는 거의 출제되지 않음

F. 분류 프레임워크와 평가 지표

분류의 일반 프레임워크

- **목표**: 데이터 포인트를 K개 클래스 중 하나에 할당 (연속값이 아님)
- 실무 예시: 이탈 예측, 광고 클릭 예측, 사기 거래 탐지

모델 유형	접근 방식	수식
Generative	X와 Y의 결합 분포를 모델링	$p(X,Y) = p(Y
Discriminative	결정 경계를 직접 학습	$y = \arg\max_k p(Y=k

- 두 방법 모두 **사후 확률을 최대화**하는 클래스를 선택함
- 이진 분류(0/1)가 기본이나, 다중 클래스로 확장 가능함

Accuracy Paradox: 왜 정확도만으로는 부족한가

- 희귀 암 진단: 10,000명 중 1명만 발생
- "모든 사람이 암이 아니다"라고 예측하면 → 정확도 **99.99%**
- 하지만 이 모델은 **완전히 무용함**

용어	정의
True Positive (TP)	양성을 정확히 양성으로 예측
False Positive (FP)	음성을 잘못 양성으로 예측 (Type I Error)
True Negative (TN)	음성을 정확히 음성으로 예측
False Negative (FN)	양성을 잘못 음성으로 예측 (Type II Error)

- 불균형 클래스에서는 Confusion Matrix 기반 지표가 필수임

Precision vs Recall Trade-off

Precision (정밀도)

$$\text{Precision} = \frac{TP}{TP + FP}$$

- "양성 예측 중 실제 양성의 비율"
- 암 진단: 양성 판정 시 **진짜 암인 비율**
- High Precision = 잘못된 양성 적음
- 하지만: 진짜 환자를 **놓칠 수** 있음

Recall (재현율)

$$\text{Recall} = \frac{TP}{TP + FN}$$

- "실제 양성 중 포착된 비율"
- 암 진단: 실제 환자 중 **감지된 비율**
- High Recall = 놓치는 환자가 적음
- 하지만: 건강한 사람을 **암으로** 오진할 수 있음

F1 Score와 ROC/AUC

- Precision과 Recall이 **동일하게 중요할 때** F1 Score 사용

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

- **ROC Curve**: 다양한 임계값에서 TPR vs FPR을 플롯
- **AUC**: ROC 곡선 아래 면적 (0~1, 높을수록 좋음)
  - 최적: 좌상단을 "감싸는" 곡선 — 높은 TPR + 낮은 FPR

지표	언제 사용하는가
Precision	FP 비용이 큰 경우 (스팸 필터 — 정상 메일 차단 방지)
Recall	FN 비용이 큰 경우 (암 진단 — 환자 놓침 방지)
F1	두 지표의 균형이 필요한 경우
AUC	전반적 분류 능력 비교 시

면접에서의 분류 지표 활용법

- 개방형 케이스 질문에서 **비즈니스 맥락**에 맞는 지표를 선택해야 함
- FP와 FN의 **비용 차이**를 명확히 설명하는 것이 핵심

시나리오	우선 지표	이유
암 진단	Recall	FN(미감지) = 환자 사망
스팸 필터	Precision	FP(오차단) = 중요 메일 손실
사기 탐지	Recall 우선	FN(미감지) = 금전 손실
추천 시스템	Precision 우선	FP(부적절 추천) = 사용자 이탈

- "False positive와 false negative의 비즈니스 영향을 설명하라" — 면접 빈출

G. 로지스틱 회귀

로지스틱 회귀: 시그모이드로 확률 변환

- 분류에서 **가장 인기 있는 알고리즘** — 선형 회귀만큼 면접에서 빈출

선형 출력을 시그모이드 함수로 확률(0~1)로 변환:

$$S(x) = \frac{1}{1 + e^{-X\beta}}$$

- $S(x) \geq 0.5$ → 클래스 "1"로 분류
- $S(x) < 0.5$ → 클래스 "0"으로 분류

$$P(Y = 1|X) = S(X\beta)$$

- 손실 함수: **Log-Loss** (Binary Cross-Entropy)
- 다중 클래스: **Softmax Regression**으로 일반화

로지스틱 회귀의 강점과 한계

강점

- **높은 해석 가능성**: 출력이 확률 → 의사결정자에게 설명 용이
- **빠른 계산 속도**: 대규모 데이터에서도 효율적
- **실무 첫 번째 모델**: 분류 문제의 baseline으로 자주 사용됨
- 선형 회귀와 유사한 직관

한계

- High Bias, Low Variance 모델 — 비선형 결정 경계에 약함
- 피처 간 높은 상관 → 계수 β의 정확도 저하
- 대응: 정규화(L1/L2), 변수 제거 등 선형 회귀와 동일한 기법 적용
- **면접 핵심**: 메커니즘 + 함정 모두 이해해야 함

H. Naive Bayes

Naive Bayes: 단순한 가정의 힘

- 적은 훈련 데이터로 빠르게 파라미터를 추정할 수 있음 → **첫 번째 모델**로 인기
- Bayes' Rule + **조건부 독립** 가정을 결합함

2가지 핵심 가정:

1. Y가 주어졌을 때 각 X_i는 다른 모든 X_j와 **독립**
2. 모든 피처에 **동일한 가중치** 부여

분류 규칙:

$$y = \arg\max_k P(Y = k) \prod_{j=1}^{n} P(X_j | Y = k)$$

Naive Bayes의 효율성과 적용 분야

- 일반적으로 k개 피처 → 2^k개 피처 상호작용 → 2^k개 데이터 필요
- 조건부 독립 가정 → **k개 데이터**만으로도 충분함

적용 분야	이유
스팸 분류	단어(피처)가 일반적으로 서로 독립적
감성 분석	텍스트 피처에 독립 가정이 자연스러움
문서 분류	고차원(많은 단어) + 독립 가정이 잘 맞음

- **현실에서 독립 가정은 거의 성립하지 않음** — 피처들은 상관되는 경향이 있음
- 그럼에도 실전에서 잘 작동함: 대부분의 데이터가 **선형 분리 가능**하기 때문

I. SVM (서포트 벡터 머신)

SVM의 핵심 아이디어

- **목표**: 훈련 데이터를 선형으로 분리하는 **초평면(hyperplane)**을 찾음
- **마진(Margin)**: 결정 경계에서 가장 가까운 훈련 포인트까지의 최소 거리
- SVM은 이 **마진을 최대화**하는 초평면을 선택함
- **서포트 벡터(Support Vectors)**: 초평면에 가장 가까운 포인트들

특성	설명
결정 경계	비선형 가능 (로지스틱 회귀와 차이점)
핵심 원리	마진 최대화 → 일반화 성능 향상
Ridge와의 관계	SVM = Ridge의 커널화(kernelized) 형태로 볼 수 있음

커널 트릭: 비선형 분리의 비밀

- 현실에서 데이터가 선형 분리 안 되는 경우가 대부분임
- **커널(Kernel)**: 데이터를 **고차원 공간으로 변환** → 그곳에서 선형 분리 가능

$$k(x, y) = \phi(x)^T \phi(y)$$

커널	사용 상황
Linear	데이터가 선형 분리 가능할 때
RBF (Radial Basis Function)	비선형 문제 — 가장 인기 있는 비선형 커널
Gaussian	RBF와 유사, 비선형 문제에 사용

- **일반 규칙**: 선형 문제 → 선형 커널, 비선형 문제 → RBF 커널

SVM: 언제 쓰고, 언제 쓰지 않는가

SVM이 좋은 경우

- **고차원 공간**: 차원 수 > 데이터 포인트 수
- 클래스 간 **명확한 초평면 분리**가 존재
- 비선형 결정 경계가 필요할 때
- 소규모 데이터셋

SVM이 안 좋은 경우

- **대규모 데이터**: 계산 복잡도가 높음
- 타겟 클래스가 겹치고 **깨끗한 분리 불가**
- **해석 가능성**이 중요한 경우 — 확률 출력이 없음
- 이 경우 로지스틱 회귀가 더 나음

SVM vs 로지스틱 회귀 vs Naive Bayes

기준	SVM	로지스틱 회귀	Naive Bayes
결정 경계	비선형 가능	선형	선형
해석 가능성	낮음	높음 (확률)	중간
대규모 데이터	느림	빠름	매우 빠름
소규모 데이터	강함	보통	강함
고차원	강함	보통	강함
출력 형태	클래스	확률	확률

- ML-heavy 역할: **커널 선택, 커널 트릭, SVM 최적화 문제**까지 알아야 함

Part 1 요약

Part 1 핵심 개념 한눈에 보기

평가/선택

- Bias-Variance Trade-off
- Overfitting 감지
- 정규화 (L1/L2)
- 해석 가능성
- 교차 검증

선형 회귀

- 4대 가정 (LINE)
- RSS, R^2, MSE, MAE
- 이분산성, QQ Plot
- 다중공선성, VIF
- 교란 변수, GLM

분류 평가

- Confusion Matrix
- Precision vs Recall
- F1 Score
- ROC/AUC
- Accuracy Paradox

고전 분류기

- 로지스틱 회귀 (시그모이드)
- Naive Bayes (조건부 독립)
- SVM (마진 최대화)
- 커널 트릭 (비선형 분리)
- 생성 vs 판별 모델

면접에서 자주 나오는 질문 TOP 7

#	질문	핵심 포인트
1	Bias-variance tradeoff란?	3가지 오차 성분 + 상황별 대응
2	Overfitting을 어떻게 감지/방지?	학습 곡선, 정규화, 교차 검증
3	L1 vs L2 차이는?	Lasso(변수 선택) vs Ridge(축소)
4	선형 회귀 가정은?	LINE: 선형성/등분산/독립/정규성
5	Precision vs Recall 차이는?	FP vs FN 비용의 비즈니스 맥락화
6	시계열에 CV를 어떻게 적용?	시간 순서 기반 forward-chaining
7	SVM vs 로지스틱 회귀 언제 사용?	비선형 경계 + 소규모 데이터 vs 해석 필요