1. 서포트 벡터 머신의 근본적인 아이디어는 무엇인가?

두 클래스를 가장 잘 구분할 수 있는 결정 경계를 찾는 것. 마진과 서포트 벡터들 사이의 거리를 최대화하는 것이 목표.

2. 서포트 벡터가 무엇인가?

결정 경계에 가장 가까이 위치한 훈련 샘플. 서포트벡터를 기준으로 마진이 정의된다.

3. SVM을 사용할 때 입력값의 스케일이 왜 중요한가?

SVM은 거리를 기반으로 한 알고리즘이기 때문에 특성의 스케일이 크면 특정 특성이 결정 경계에 지나치게 큰 영향을 미치게 된다. 따라서 입력값을 표준화하여 스케일을 맞춰주는 것이 매우 중요하다.

4. SVM 분류기가 샘플을 분류할 때 신뢰도 점수와 확률을 출력할 수 있는가?

신뢰도 점수는 기본적으로 제공되며, 이는 초평면으로부터의 거리이다. 확률값은 기본적으로 제공되지 않지만 platt scaling을 통해 확률로 변환된 값을 제공받을 수 있다.

5. RBF 커널을 사용해 SVM 분류기를 훈련시켰더니 훈련 세트에 과소적합되었다. 이때, 어떤 값을 어떻게 수정해야 하는가?

과소적합은 모델의 복잡도가 너무 낮을 때 발생한다. c값을 증가시켜 마진 오류에 대한 패널티를 강화하거나 gamma 값을 증가시켜 RBF 커널 영향 범위를 좁힌다.