

Curse of dimensionality به پدیده‌ای اشاره دارد که با افزایش تعداد ابعاد یا همان ویژگی‌ها در داده‌ها، حجم فضای ویژگی به صورت نمایی افزایش می‌یابد. این باعث می‌شود:

1. داده‌ها در فضای با ابعاد بالا بسیار پراکنده شوند

2. نیاز به نمونه‌های آموزشی به صورت نمایی افزایش یابد

علاوه بر این‌ها مشکلات دیگری هم در خود SVM به وجود می‌آید مثل:

1. مشکلات در کرنل

SVM اغلب از کرنل برای نگاشت داده‌ها به فضای با ابعاد بالاتر استفاده می‌کند تا داده‌های غیرخطی‌پذیر را جدا کند. اما در ابعاد بسیار بالا:

- محاسبات هسته پیچیده و پرهزینه می‌شود

- خطر بیش‌برازش (overfitting) افزایش می‌یابد

2. کاهش کارایی طبقه‌بندی

در ابعاد بالا:

- مرز تصمیم‌گیری ممکن است بیش از حد پیچیده شود

- عملکرد تعمیم‌پذیری مدل کاهش می‌یابد یعنی ممکن است در یک دیتاست دیگر خطای زیادی ایجاد کند

3. مشکلات محاسباتی

- پیچیدگی محاسباتی SVM با افزایش ابعاد به شدت افزایش می‌یابد

- نیاز به حافظه بیشتر برای ذخیره ماتریس کرنل

- زمان آموزش مدل به صورت نمایی رشد می‌کند

ریشه اصلی اکثر این مشکلات به دلیل این است که ویژگی‌های بی‌اهمیتی وارد دستگاه شده یا اینکه دیتاست ما کوچک است. به هر حال اگر ویژگی‌های انتخاب شده به اندازه کافی مهم باشند همانطور که قبلاً گفته شد دیتاست باید به تعداد حالات آن ویژگی چند برابر شود و در غیر این صورت اگر ویژگی تأثیر کمتری دارد اصلاً باید به شیوه‌ای حذف شود. برای حل این مشکلات از این روش‌ها می‌توان استفاده کرد:

1. Feature Selection

استفاده از روش‌هایی مانند:

- فیلترهای آماری (مانند آزمون  $t$  یا آنالیز واریانس)
- روش‌های مبتنی بر مدل (مانند اهمیت ویژگی در درختان تصمیم)
- الگوریتم‌های جستجو مانند SFS یا SBS

## 2. Dimensionality Reduction

- PCA (تحلیل مؤلفه‌های اصلی)
- LDA (تحلیل تشخیصی خطی)
- روش‌های مبتنی بر یادگیری مانیفولد مانند t-SNE یا UMAP

## 3. روش‌های نمونه‌برداری

- افزایش حجم داده‌های آموزشی متناسب با افزایش ابعاد
- استفاده از روش‌های نمونه‌برداری هوشمند