



## ۱ انتخاب ویژگی

انتخاب ویژگی را با عناوین subset selection / feature selection می‌شناسیم. اگر بخواهیم بهترین زیرمجموعه از ویژگی‌ها را انتخاب کنیم عملاً باید  $2^p$  زیرمجموعه ممکن را آموزش دهیم و از بین آنها بهترین را که از دیدگاه اعتبارسنجی برتر است را انتخاب نماییم. به این روش best subset selection می‌گوییم. عملاً اگر تعداد ویژگی‌ها از ۴۰ بیشتر باشد چنین محاسباتی غیرممکن است و باید از روش‌های محاسباتی زیربهبوده استفاده کنیم.

برای مثال برای  $p = 20$  داریم:

$$2^p = 1,048,576$$

درحالی‌که:

$$1 + \frac{p(p+1)}{2} = 211$$

یکی از این روش‌ها انتخاب قدم به قدم می‌باشد.

### ۱.۱ قدم به قدم پیش رونده Forward Stepwise Selection

در این روش ابتدا یکی از ویژگی‌ها را که بهترین کیفیت را دارد انتخاب کرده و در قدم بعدی ویژگی جدیدی اضافه می‌کنیم که بهترین عملکرد را دارد. اگر در مرحله اول هیچ ویژگی را وارد نکنیم، آنگاه در مجموع

$$1 + \sum_{k=0}^{p-1} (p-k) = 1 + \frac{p(p+1)}{2}$$

مدل آموزش داده شده که بسیار کمتر از  $2^p$  برای یافتن بهترین زیرمجموعه می‌باشد.

### ۲.۱ قدم به قدم پس رونده Backward Stepwise Selection

در این حالت از یک مدل کامل شروع می‌کنیم و ویژگی‌ها را یکی یکی کم می‌کنیم. در آخر بین مدل‌های انتخاب شده آنرا که در اعتبارسنجی به عنوان مدل برتر انتخاب شده است، را انتخاب می‌کنیم.

### ۳.۱ ترکیبی Hybrid

در این روش به عنوان مثال پیش رونده می‌رویم؛ ولی اگر ویژگی وجود داشته باشد که کیفیت را بهتر نمی‌کند آنرا حذف می‌کنیم.

### ۴.۱ روش تحلیلی

برای انتخاب مدل برتر از روش تحلیلی نیز می‌توانیم استفاده کنیم. معیارهای زیر در عمل مورد استفاده قرار می‌گیرند.

$$C_p = \frac{1}{n}(RSS + 2d\hat{\sigma}^2) \quad ۱.$$

$$AIC = C_p \quad ۲.$$

$$BIC = \frac{1}{n}(RSS + \log nd\hat{\sigma}^2) \quad ۳.$$

$$Adjusted R^2 = 1 - \frac{\frac{RSS}{n-d-1}}{\frac{TSS}{n-1}} \quad ۴.$$

AIC و BIC را می‌توان در موارد غیر رگرسیون خطی نیز مورد استفاده قرار داد.