Number of Terms for Polynomial Regression

K.M. Chugg, B.K. Jenkins
version 0.3

March 11, 2023

1 Polynomial Vectors for Nonlinear Models

Consider a $D \times 1$ feature vector $\underline{x} = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_D \end{bmatrix}^T$. For linear regression/classification, we augment this with a zero-power term to obtain $\underline{x}^{(+)} = \begin{bmatrix} 1 & x_1 & x_2 & x_3 & \cdots & x_D \end{bmatrix}^T$. For polynomial regression/classification, the vector comprising this augmented feature and higher order cross terms of the components is considered. For the case of $D = 3$ and second order terms, we have

$$\underline{\phi}_2(\underline{x}) = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \\ x_1^2 \\ x_1 x_2 \\ x_1 x_3 \\ x_2^2 \\ x_2 x_3 \\ x_3^2 \end{bmatrix} \tag{1}$$

Thus, the original 3 dimensional feature vector maps to a 10 dimensional vector considering all powers up to $P = 2$. Using larger values of P will provide a more powerful model, but the dimension of $\underline{\phi}$ also grows substantially with P.

1.1 Dimension for D Features and Highest Degree P

Let's consider how quickly this growth occurs with the parameters D and P. Consider the vector $\underline{\phi}_P(\underline{x})$ that contains all terms up to power P – *e.g.,* the example in (1) has $P = 2$ and $D = 3$. Denote the dimension of $\underline{\phi}_P(\underline{x})$ by $K(D, P)$. We know that $K(D, 1) = D + 1$ because there is one term of power 0 and D terms of power 1 – *i.e.,* $\underline{\phi}_1(\underline{x}) = \underline{x}^{(+)}$. To find an expression for $K(D, P)$, we can first consider how many terms there are of a given power p with a fixed value of D. From the example above with $D = 3$, there are 6 terms with $p = 2$, 3 terms with $p = 1$, and 1 term with $p = 0$ so that $K(D = 3, P = 2) = 6 + 3 + 1 = 10$. More generally, the a term with power p with D features is of the form

$$x_1^{m_1} x_2^{m_2} \cdots x_D^{m_D} \tag{2}$$

satisfying

$$m_1 + m_2 + \cdots + m_D = p \tag{3}$$

where each m_i is an non-negative integer. The number of solutions to (3) is called the *occupancy number* and it is a well known problem in combinatorics. In fact, this is the number of ways to draw p objects from D without order and with replacement. Specifically, the number of solutions to (3) is

$$K_p(D) = \binom{D+p-1}{p} \tag{4}$$

For example, when $D = 3$, $K_2(3) = \binom{3+2-1}{2} = \binom{4}{2} = 6$ which we already know through the enumeration in (1). See Section 1.2 below if you are curious how to arrive at (4) or if your memory of combinatorics is a bit rusty.

With the result in (4) we can now compute the the dimension of $\underline{\phi}_P(\underline{x})$

$$K(D,P) = \sum_{p=0}^{P} K_p(D) = \sum_{p=0}^{P} \binom{D+p-1}{p} = \sum_{p=0}^{P} \frac{(D+p-1)!}{(D-1)!p!} \tag{5}$$

There is a simpler expression for $K(D,P)$ that can be derived using properties of the binomial coefficient

$$\boxed{K(D,P) = \binom{D+P}{P}} \tag{6}$$

See Section 1.3 for the details. As an example, for a quadratic ($P = 2$), there are

$$K(D,2) = \binom{D+2}{2} = \frac{(D+2)(D+1)}{2} = \frac{D^2+3D+2}{2} = \frac{1}{2}D^2 + \frac{3}{2}D + 1 \tag{7}$$

This expression for $K(D,P)$ in (6) is plotted in Figure 1 for $D \in \{2,3,4,5\}$ and $P \leq 4$. The number of terms grows quite rapidly. To consider larger values of D and P, a log-scale is used to plot $K(D,P)$ in Figure 2, Can you determine the asymptotic growth rate of $K(D,P)$ in P for a fixed D? How about the asymptotic growth rate of $K(D,P)$ in D for a fixed P? Even without computing these explicitly, it is clear that the dimension of $\underline{\phi}_P(\underline{x})$ grows very rapidly in both P and D and therefore only small values of P will be useful in practice.

1.2 Derivation of the Occupancy Number

First recall a result from combinatorics: given n objects that are to be partitioned into two groups, one of size k and the other of size $n-k$ with $0 \leq k \leq n$, then there are $\binom{n}{k}$ ways to perform this partition. For example, if you have 20 people and need to chose a team of 5, there are $\binom{20}{5} = \binom{20}{15}$ ways to form a 5 person team. This can be seen as partitioning the 20 people into two groups (*i.e.,* "on team" and "not on team").

Now consider the number of solutions to (3). For concreteness, consider $p = 6$ and $D = 3$ so that

$$m_1 + m_2 + m_3 = 6 \tag{8}$$

with each m_i a non-negative integer. A specific solution to this is $m_1 = 2$, $m_2 = 1$ and $m_3 = 3$ which can be represented by

$$\checkmark\checkmark + \checkmark + \checkmark\checkmark\checkmark = 6 \tag{9}$$

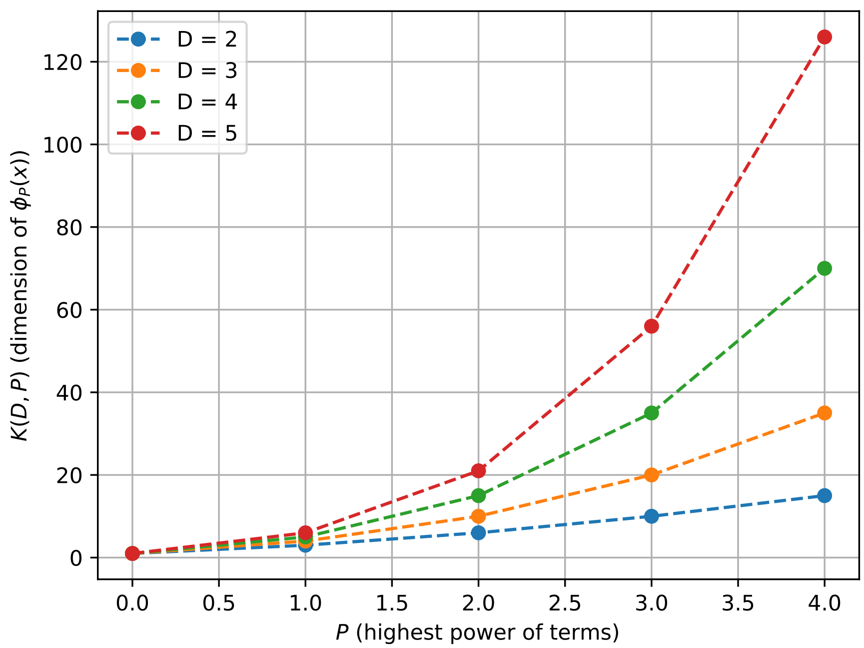

Figure 1: Dimension of polynomial vector with D features.

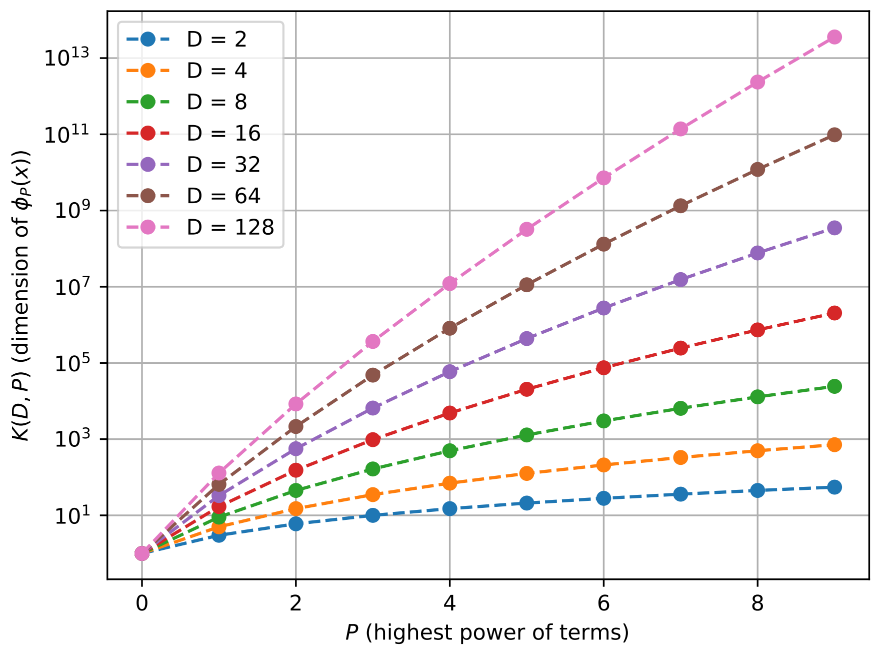

Figure 2: Dimension of polynomial vector with D features with logarithmic y-axis.

where we have replaced the non-negative integer m_i by the corresponding number of check marks. Similarly, the $m_1 = 2$, $m_2 = 0$ and $m_3 = 4$ solution can be represented by

$$\checkmark\checkmark + + \checkmark\checkmark\checkmark\checkmark = 6 \tag{10}$$

Notice that if we use this representation in (3), there will be $D-1$ plus signs ($+$) and p checkmarks ($\checkmark$). Thus, there are $D-1+p$ symbols in total on the left-hand side of (3). This can be thought of as $D-1+p$ placeholder positions that must be assigned a $+$ or $\checkmark$. Thus, this is a partition of $D-1+p$ objects into two groups, one of size $D-1$ and the other of size p. Using the above fact about binary partitions yields the occupancy number in (4).

1.3 Derivation of the Required Binomial Sum Identity

To go from (5) to (6) we use the following identity from combinatoriics

$$\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k} \tag{11}$$

This is simple to verify algebraically, but it also has a simple combinatoric proof. Specifically, consider a room with n students and 1 teacher. It is desired to send k people outside of the room and keep $n+1-k$ inside the room. This is a binary partitioning problem and the number of possibilities is $\binom{n+1}{k}$, which is the right hand side of (11). We can interpret the left hand side of (11) by considering the teacher separately. If the teacher is held in the room, then we need to pick k from n students to go outside, for which there are $\binom{n}{k}$ possibilities. If the teacher is taken outside the room, then we need to pick $k-1$ from n students to go outside, for which there are $\binom{n}{k-1}$ possibilities. Thus, by conditioning on the teach inside or outside the room, we arrive at (11).

This can be applied repeatedly to obtain

$$\sum_{k=0}^{r} \binom{n+k-1}{k} = \binom{n+r}{r} \tag{12}$$

which is the identity used to go from (5) to (5). To show (12), we can use induction and apply (11). Equivalently, we can repeat the combinatorial argument used to arrive at (11), but now with n students and r teachers and a goal of placing r people outside the room. You can label the teachers 1 through r and first consider whether teacher 1 is inside/outside the room yielding the equivalent of (11)

$$\binom{n+r-1}{r-1} + \binom{n+r-1}{r} = \binom{n+r}{r} \tag{13}$$

Then apply the same argument to teacher 2 to the $\binom{n+r-1}{r-1}$ in (13). Repeating this r times yields (12).

In fact, this argument provides a more direct path to (6). In this context, D corresponds to the number of students, P to the number of teachers, and P to the number of people taken outside of the room. This is not easy to see at first glance, however, and therefore a two-step argument is applied herein.