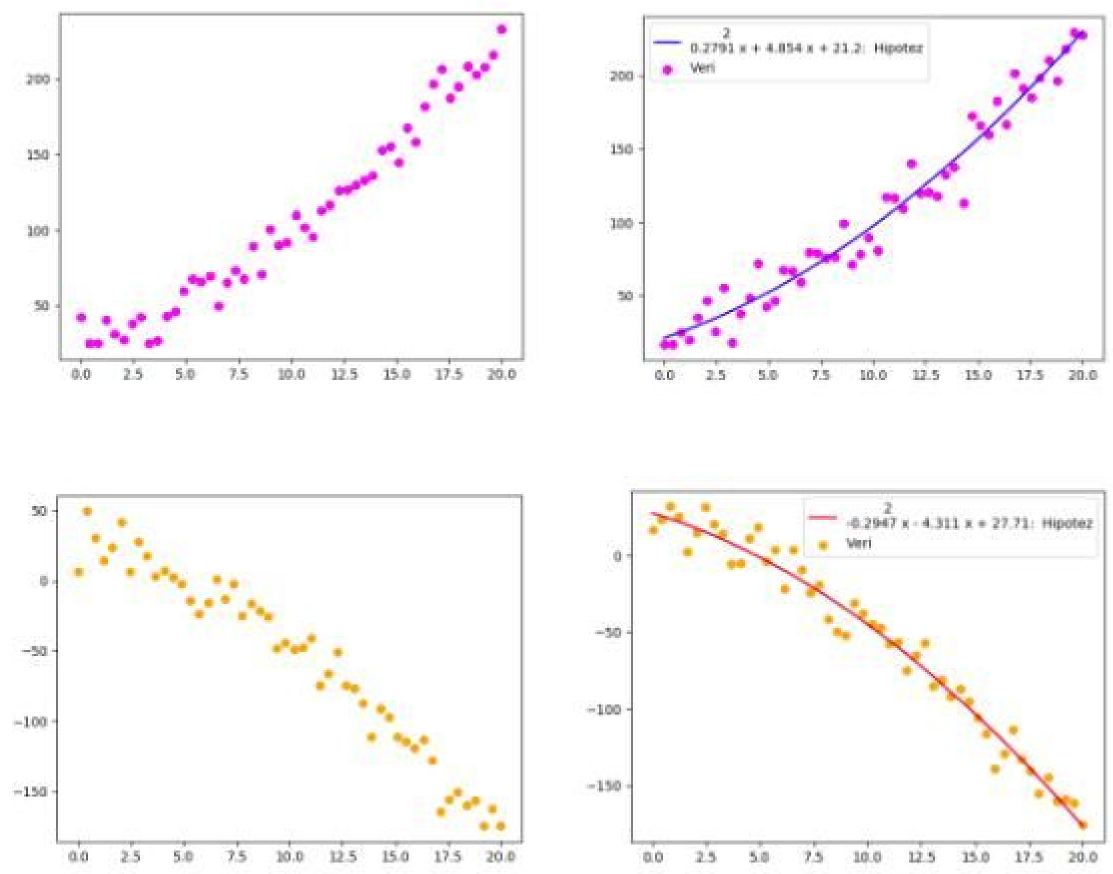
3.5 Polinomsal Regresyon

Artık hipotezimizi nasıl geliştirebileceğimize bakabiliriz. Regresyon problemleri her zaman doğrusal olmak zorunda değildir. Yani çizdiğimiz eğri her zaman bir doğru denklemi ile ifade edilemeyebilir. Böyle durumlara polinomsal regresyon denir.

- Polinomsal Regresyon: Veriler arasında doğrusal olmayan bir ilişki olduğunda, hipotez fonksiyonumuzun davranışını veya eğrisini, kuadratik, kübik veya karekök fonksiyon (veya başka herhangi bir form) yaparak değiştirebiliriz.

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \ldots + \beta_h X^h + \epsilon,$$

Aşağıda bazı verilerin dağılım grafikleri verilmiş. Bu verilere baktığımızda doğrusal lineer regresyon uygulayabileceğimizi görebilirsiniz. Ancak Polinomsal regresyonun uygulanmış haline bakarsanız verilere daha iyi fit ettiğini (uyum sağladığını) görebilirsiniz.



Hipotez fonksiyonumuzun özelliklerini ve biçimini birkaç farklı yoldan geliştirebiliriz. Birden çok özelliği birleştirebiliriz. Örneğin, x1 ve x2, x1 · x2 alarak yeni bir özellik x3 eklenebilir. Hipotez işlevi, verilere iyi uymuyorsa doğrusal (düz çizgi) olmasına gerek yoktur. Hipotez fonksiyonumuzun davranışını veya eğrisini, kuadratik, kübik veya karekök fonksiyon (veya başka herhangi bir form) yaparak değiştirebiliriz. Bazen, veriler arasında doğrusal olmayan bir ilişki olabilir, böyle bir ilişkiyi açıklamaya çalışmanın bir yolu, bir polinom regresyon modelidir. Tek bir tahmini X için böyle bir model:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \ldots + \beta_h X^h + \epsilon,$$

Burada h polinom derecesi olarak adlandırılır. Daha düşük dereceler için, ilişki belirli bir ada sahiptir (yani, h = 2'ye kuadratik, h = 3'e kübik, h = 4'e kuartik denir vb.). Bu model, Y ve X arasındaki doğrusal olmayan bir ilişkiyi mümkün kılmasına rağmen, polinom regresyonu, regresyon katsayıları β1, β2, …, βh'de doğrusal olduğu için yine de doğrusal regresyon olarak kabul edilmektedir.

Yukarıdaki denklemi hesaplamak için, yalnızca yanıt değişkenine (Y) ve öngördürücü değişkenine (X) ihtiyacımız olacaktır. Bununla birlikte, polinom regresyon modelleri de etkileşim terimine yol açabilecek diğer tahmin değişkenlerine sahip olabilir. Gördüğünüz gibi, yukarıdaki bir polinom regresyon modeli için temel denklem nispeten basit bir modeldir ancak durumunuza bağlı olarak modelin nasıl büyüdüğünü hayal edebilirsiniz.

İkinci derece polinom modelinin matrisleri şöyledir:

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{50} \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_{50} & x_{50}^2 \end{pmatrix}, \vec{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}, \vec{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_{50} \end{pmatrix}$$

Burada Y ve X'deki girdiler ham verilerden oluşur. Gördüğünüz gibi, çoklu doğrusal regrasyonda kullanılan analiz tekniklerinin (ör. OLS) burada uygulanabilir.

Bir polinom regresyon modeli tahmininde akılda tutulması gereken bazı genel kurallar şunlardır:

- Fit eden model, daha büyük bir numune boyutu üzerine kurulduğunda daha güvenilirdir.
- Gözlemlenen değerlerinizin sınırlarının ötesinde, özellikle de polinom fonksiyonunda belirgin bir eğri olduğu zaman, bir ekstrapolasyonun modelin kapsamının ötesinde anlamsız sonuçlar ürettiği durumlarda ekstrapolasyon yapmayın.
- Kullanılan istatistiksel yazılım için sayısal taşmalara neden olabileceğinden yüksek dereceli terimlerin eklenmesinde öngörücünün / ölçütlerin ne kadar büyük olacağını düşünün.
- Daha yüksek dereceli bir terim eklemek için kesinlikle düşük p-değerleriyle gitmeyin, aksine yalnızca elde edilen kalıntı plotlar makul görünüyorsa modelinizi desteklemek için kullanın. Bu, "istatistiksel önem" karşısında "pratik önem" belirlemeniz gereken bir durumun bir örneğidir.
- Genel olarak, regresyon modellemesi boyunca standart uygulamalar olduğu gibi, modelleriniz modelinizin Xh'yi içeriyorsa ve Xh'nin Y'nin istatistiksel olarak önemli bir öngörücüsü olduğu gösterilirse, modelinizin her bir Xj'yi de

içermesi gerektiğini söyleyen hiyerarşi ilkesine uymanız gerekir. Tümü j <h, bu alt sınıra ait terimlerin katsayılarının önemli olup olmamasına bakılmaksızın.