6.4 Polynomické regresní modely

Při hledání optimálního stupně m polynomu (str. 372 v cit.[72])

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \dots \beta_m x^m \quad ,$$

kde β_0, β_1, β_2, β_3, ..., β_m, jsou odhadované parametry, patří mezi často užívaná rozhodčí kritéria *střední kvadratická chyba predikce*

$$MEP = \frac{1}{n} \sum_{i=1}^{n} (y_i - \boldsymbol{x}_i^{\mathbf{T}} \boldsymbol{b}_{(i)})^2 \quad ,$$

kde $\boldsymbol{b}_{(i)}$ je odhad parametrů regresního modelu určený ze všech bodů kromě i-tého a $\boldsymbol{x}_i$ je i-tý řádek matice $\boldsymbol{X}$. Statistika *MEP* využívá predikce $\hat{y}_{P,i}$ z odhadu, při jehož konstrukci byla informace o i-tém bodu vypuštěna. Dá se snadno ověřit, že pro *MEP* platí vztah

$$MEP = \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{e}_i^2}{(1 - H_{ii})^2} \quad ,$$

Pro velké rozsahy dat n jsou prvky $H_{ii} \approx 0$ a pak $MEP = RSC/n$, dle str. 374 v cit[72]. Užije-li se charakteristiky *MEP* místo *RSC* ve výpočtu koeficientu determinace, bude výsledkem *predikovaný koeficient determinace* D_P značený v literatuře také R^2_P

$$\hat{D}_P = 1 - \frac{n\ MEP}{\sum_{i=1}^{n} y_i^2 - n\bar{y}} = \hat{R}_P^2 \quad ,$$

Univerzální použití mají také rozličná kritéria vycházející z teorie informace a entropie. Mezi nejznámější patří *Akaikovo informační kritérium AIC*

$$AIC = n \ln\left(\frac{RSC}{n}\right) + 2m \quad .$$

Při hledání stupně polynomu je za nejvhodnější považován takový model, pro který je *AIC*, ale také *MEP* minimální, zatímco D_P maximální.