P5

From slides 28 and 42 in lecture 4's notes we learnt that

$$f(x) = \sum_{k=1}^{K} \tilde{\theta}_k N_k(x)$$ for a natural cubic spline with K knots and K basis functions. N_k for $k = 1, 2, \ldots K$ are the K basis functions

$$\hat{\theta} = (N^T N + \lambda R)^{-1} N^T Y$$

$$\hat{f}(x) = \sum_{k=1}^{K} \hat{\theta}_k N_k(x)$$

Therefore we can easily compute $(N^T N + \lambda R)^{-1} N^T$ and subsequently the variance of $\hat{\theta}$, and that for $\hat{Y}$. However, UnivariateSpline's get_knots() and get_coeffs() give us K coefficients and $K-2$ knots respectively, and so I'm not completely sure what basis functions it used because to compute our N_ks we'd need as many knots as there are coeffs. I later found online that this UnivariateSpline uses some B-spline as basis functions under the hood but I didn4 find how specifically they computed it (and there also seemed to be some more parameters into certain functions that I'd need to specify), but anyways, we'll omit the computer plotting for now but instead provide a hand-written formula for the confidence interval of $\hat{Y}$, which is

$$\hat{f}(x) \pm 2 \times \sqrt{N(x)^T \operatorname{Var}(\hat{\theta}) N(x)}$$

just like for polynomial regression,

$$CI\ \hat{Y} = P(x)^T \hat{\beta} \pm 2 \times \hat{\sigma} \sqrt{P_{(x)}^T \left(P(X)^T P(X)\right)^{-1} P_{(x)}}$$

Hope you can only deduct very little point due to the difficulty and

Capital X

tedium of computing the B-splines correctly, especally when we are not given their formulae and don't know what functions to use or what parameters / configuration to set.

Please then open and review p5.ipynb. Thank you!