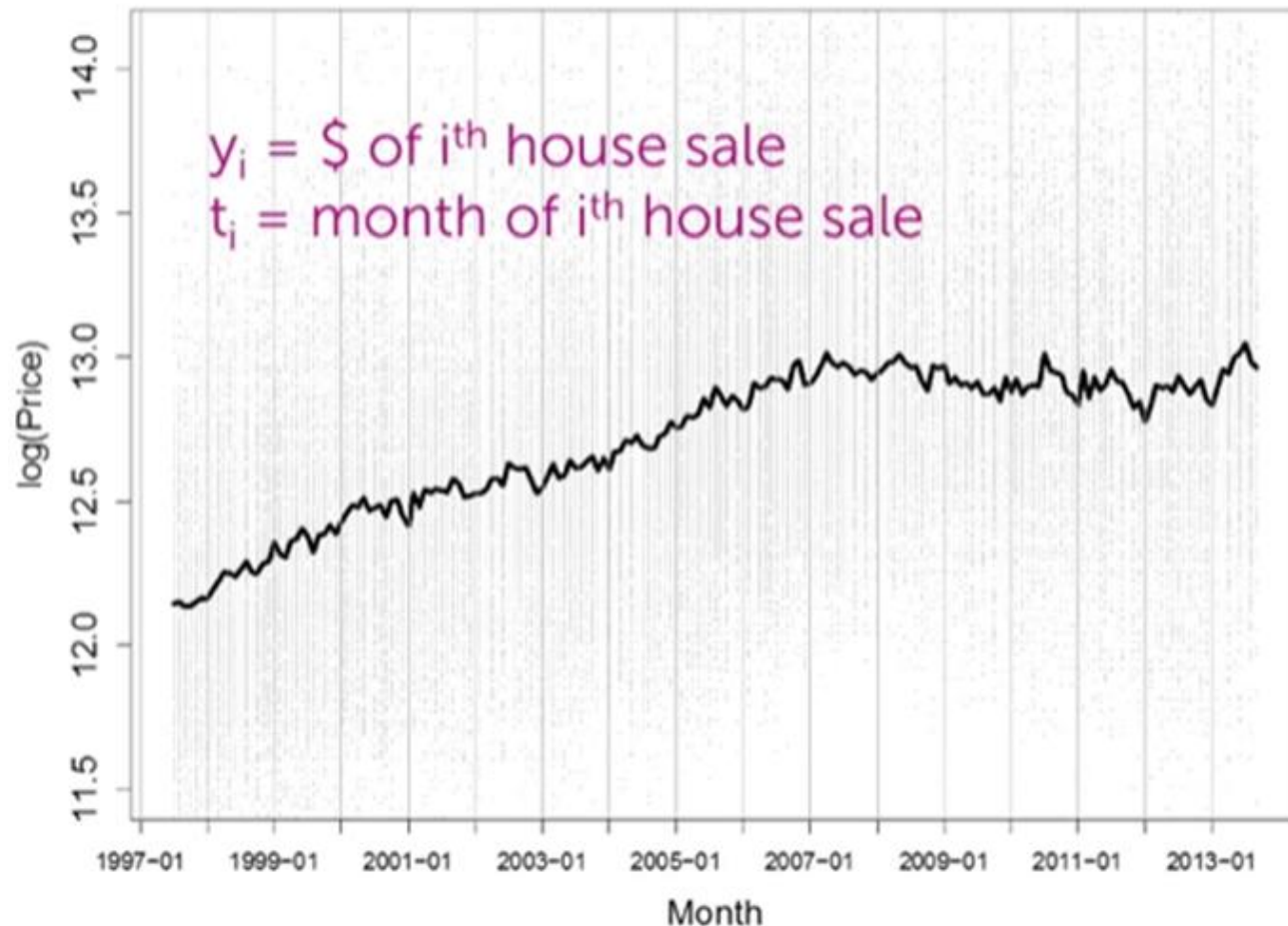


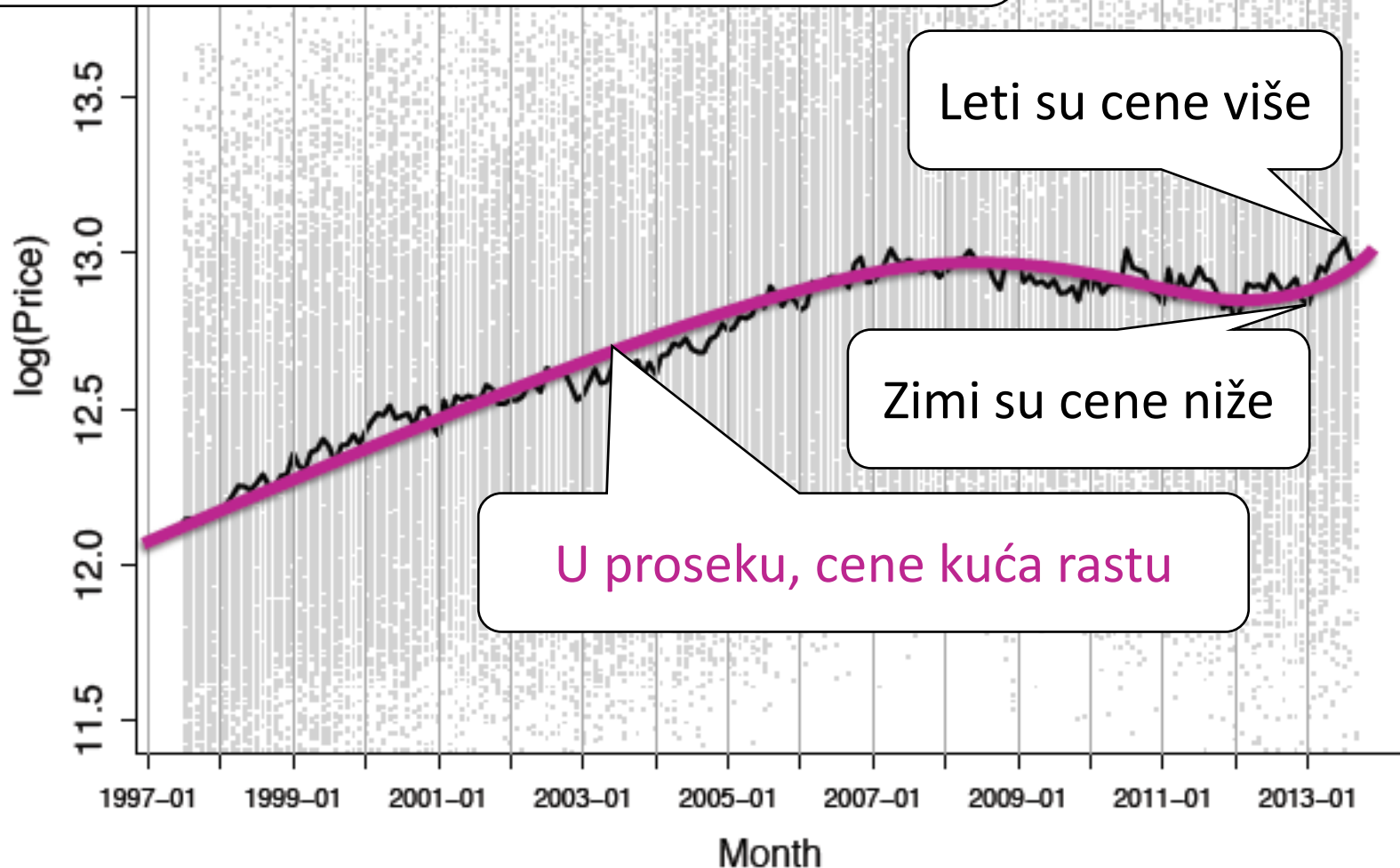
Drugi tipovi obeležja – sezonalnost

- Motivacioni primer: posmatranje trenda prosečne cene nekretnina kroz vreme (vremenske serije)



Drugi tipovi obeležja – sezonalnost

Sezonalnost – šablon koji se ponavlja
(ovde oscilacije cena na mesečnom nivou)



Drugi tipovi obeležja – sezonalnost

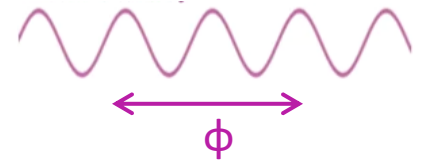
$$w_0 + w_1 t + w_2 \sin\left(\frac{2\pi t_i}{12} - \Phi\right) + \varepsilon_i$$

Linearni trend:
modeluje rast
sa vremenom



Sinusoidalna
komponenta sa periodom
12 (resetuje se godišnje)

Nepoznat pomera



Parametar Φ je unutar
sinusa – nije linearna
regresija!

Drugi tipovi obeležja – sezonalnost

$$\sin(a - b) = \sin(a) \cos(b) - \cos(a) \sin(b)$$

- Model ekvivalentno možemo predstaviti na sledeći način:

$$w_0 + w_1 t + w_2 \sin\left(\frac{2\pi t_i}{12}\right) + w_3 \cos\left(\frac{2\pi t_i}{12}\right) + \varepsilon_i$$

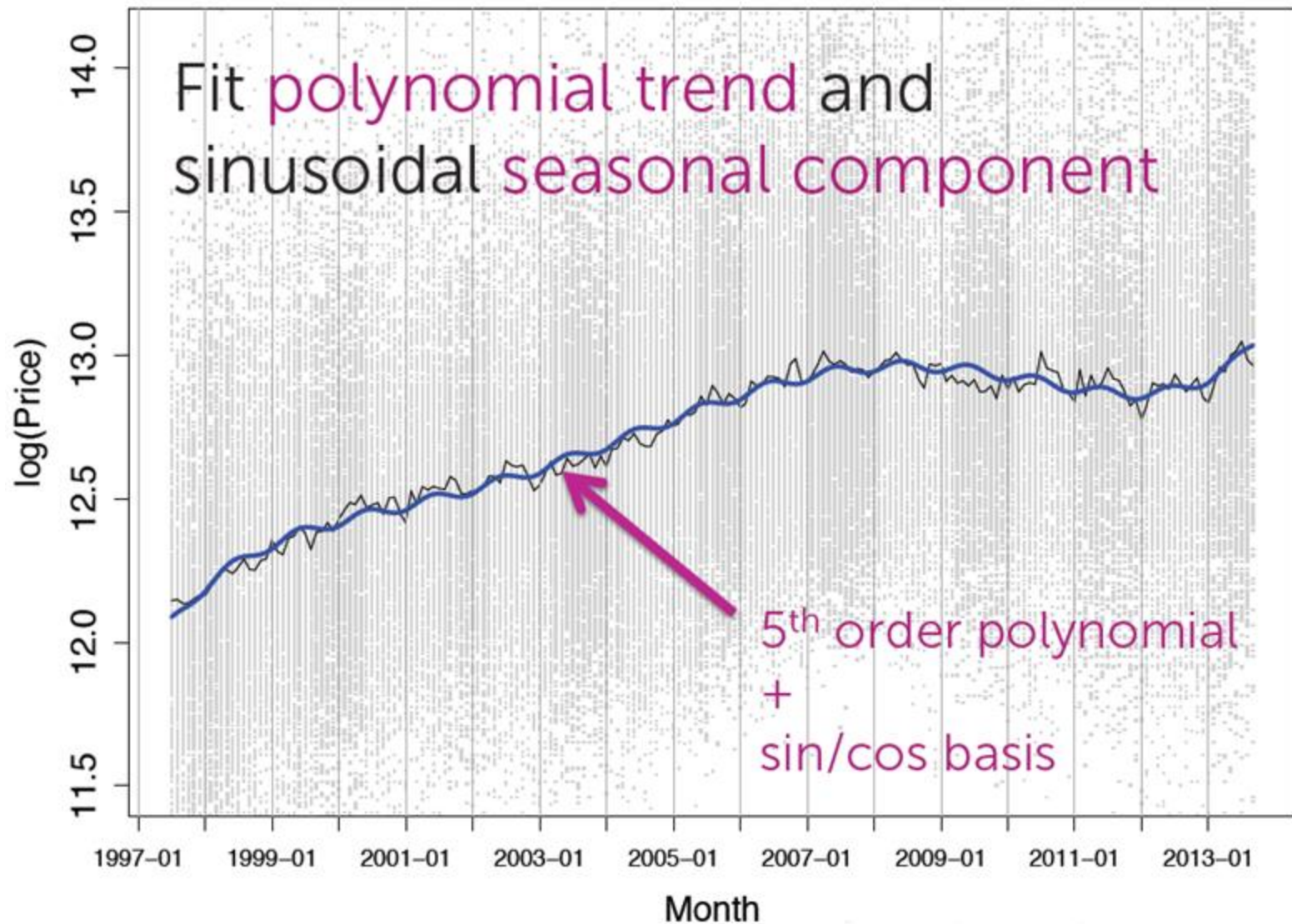
Obeležje 1: konstanta

Obeležje 2: t

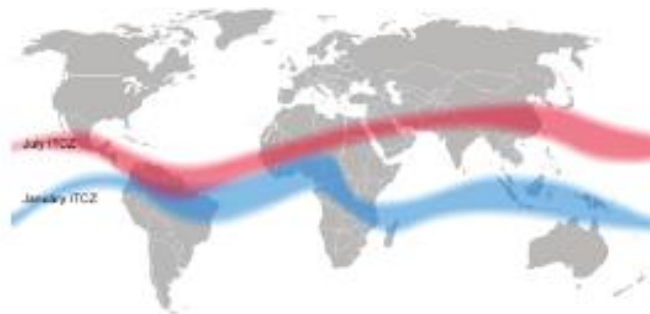
Obeležje 3: $\sin\left(\frac{2\pi t_i}{12}\right)$

Obeležje 4: $\cos\left(\frac{2\pi t_i}{12}\right)$

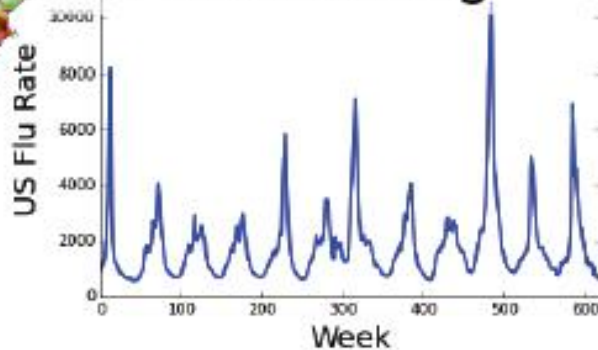
Drugi tipovi obeležja – sezonalnost



Weather modeling (e.g., temperature, rainfall)



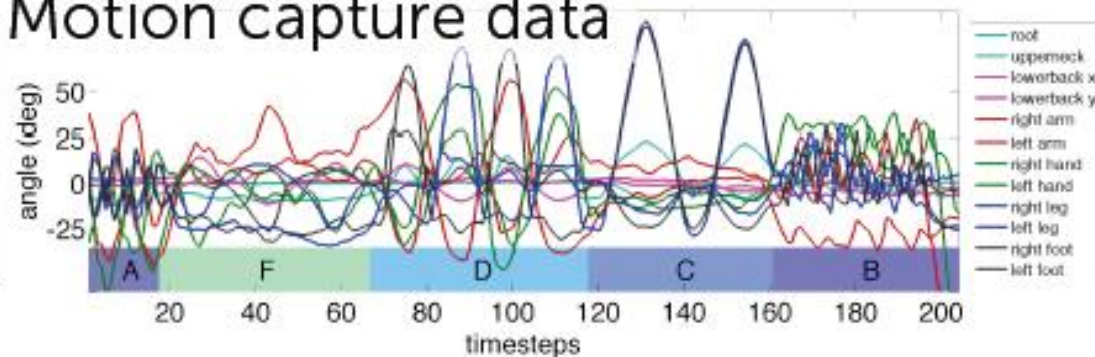
Flu monitoring



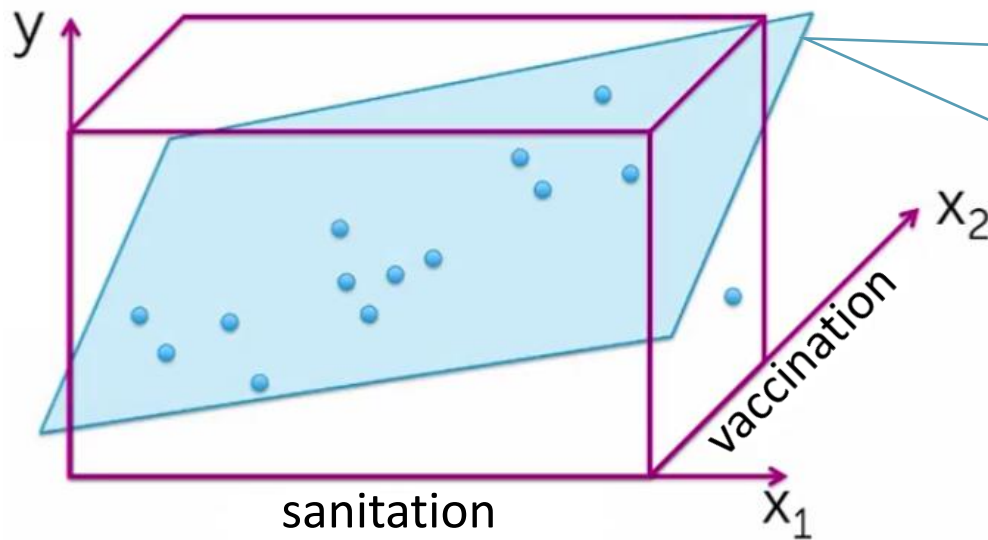
Demand forecasting (e.g., jacket purchases)



Motion capture data



Dodavanje više ulaznih varijabli



$$\begin{aligned} h_{\theta}(x) \\ &= \theta_0 + \theta_1 \cdot \textit{sanitation} \\ &\quad + \theta_2 \cdot \textit{vaccination} \end{aligned}$$

$$h_{\theta}(x) = \sum_{j=0}^d \theta_j f_j(x)$$

- obeležje 1: $f_0(x) = x^0 = 1$ (konstanta)
- obeležje 2: $f_1(x) = \textit{sanitation}$
- obeležje 3: $f_2(x) = \textit{sanitation} * \log(\textit{vaccination})$
- ...

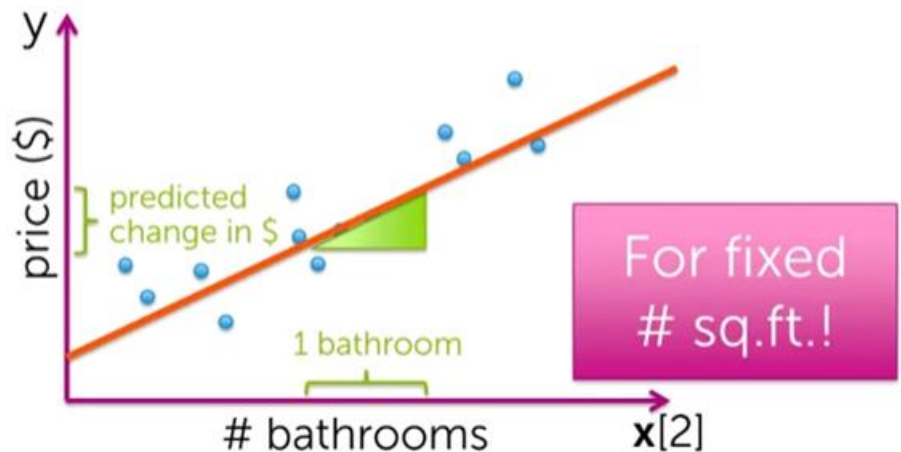
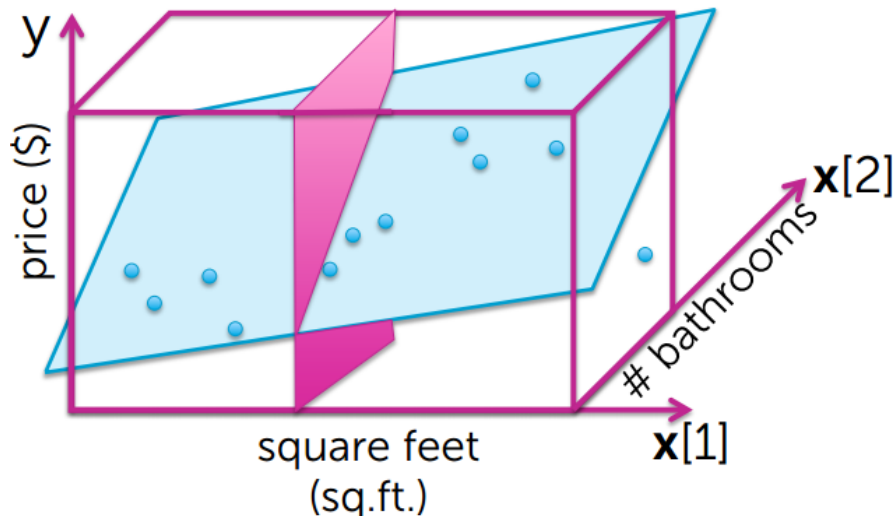
Interpretacija parametara modela θ

Kada interpretiramo koeficijent θ_j koji stoji uz obeležje x_j moramo to raditi u kontekstu svega što je uključeno u model

Moramo fiksirati

Želimo da interpretiramo

$$h_{\theta}(x) = \theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2$$



Interpretacija parametara modela θ

$$\text{cena} = \theta_0 + \theta_1 \cdot \text{kvadratura} + \theta_2 \cdot (\text{br. spavaćih soba})$$

Kako to da interpretiramo $\theta_2 < 0$?

Ako je kvadratura fiksirana,
povećanje broja spavaćih soba
znači da su one manje

Dodavanje spavaćih soba
može smanjiti cenu kuće ako
ljudi preferiraju veće sobe

Interpretacija parametara modela θ

Iz modela smo uklonili kvadraturu (i fitovali na iste podatke)

$$\text{cena} = \theta_0 + \theta_1 \cdot (\text{br. spavaćih soba})$$

Ova promena može
rezultovati sa $\theta_1 > 0$

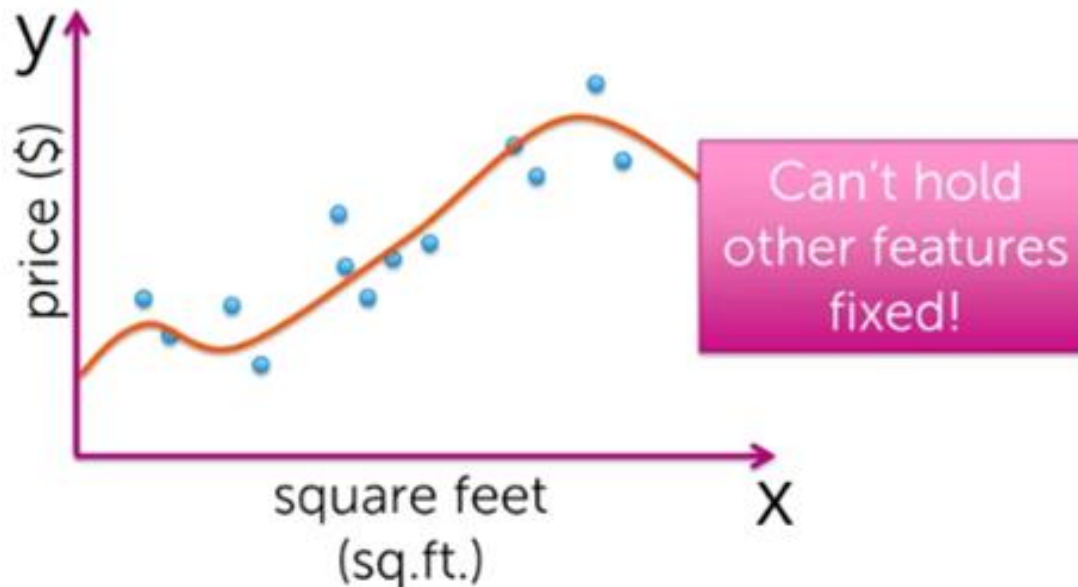
Veći broj spavaćih soba je
indikacija da kuća ima veću
kvadraturu

Interpretacija parametara modela θ

Šta je sa interpretacijom θ kod polinomijalne regresije?

$$h_{\theta}(x) = \sum_{j=0}^d \theta_j x^j$$

kako da interpretiramo
 θ_j uz x^j ?



Dodavanje ulaznih varijabli

- Jedna ulazna varijabla:
 - Možemo uvećati fleksibilnost modela dodavanjem viših stepena polinoma
 - Možemo dodavati i drugačije vrste obeležja (npr. sinusne funkcije)
- Možemo dodavati više nezavisnih ulaznih varijabli
 - npr. pored sanitarija možemo kao ulaz koristiti i % saobraćajnih nesreća, BMI, GDP, konzumacija duvana i alkohola, ishrana,...
- Možemo kombinovati ova dva pristupa

Dodavanje ulaznih varijabli

- Dodavanje obeležja uvećava kompleksnost modela
- Modeli niske kompleksnosti mogu biti nedovoljno fleksibilni da opišu podatke
 - Veliko sistematsko odstupanje (*bias*), mala varijansa
- Kompleksni modeli su fleksibilniji ali mogu da pate od *overfittinga*
 - Velika varijansa, a malo sistematsko odstupanje

Dodavanje ulaznih varijabli

- Overfitting se dešava u situacijama gde imamo mnogo obeležja (veliko D), a premalo trening podataka (malo N)
- Jedna od indikacija overfitinga jeste velika magnituda koeficijenata modela θ