

SANNOLIKHETSLÄRA OCH STATISTIK

FÖRELÄSNING 12

Mattias Villani

**Avdelningen för Statistik och Maskininlärning
Institutionen för datavetenskap
Linköpings universitet**



ÖVERSIKT

- ▶ Prediktion
- ▶ Beslut

PREDIKTION

- ▶ **Prediktion:** inferens för okända men potentiellt observerbara kvantiteter.
 - ▶ antalet buggar i en kod.
 - ▶ Sjukdom.
 - ▶ Mängden brytbara mineraler.
- ▶ **Prognos:** prediktion av framtida utfall.
 - ▶ lägenhetspriserna i linköping Jan 2016.
 - ▶ Framtida försäljning.
 - ▶ Slutpris i en eBay auktion.

PREDIKTION FÖR ATT UTVÄRDERA MODELLER

- ▶ Prediktion är också ett utmärkt sätt att **utvärdera** och **jämföra modeller**. En korrekt modell predikterar bra.
- ▶ **Träningsdata** - **testdata**.
- ▶ **Accuracy**: antal korrekt klassificeringar / totala antalet test data observationer.
- ▶ Handwritten digits data (3000 obs för träning, 10000 för test):
 - ▶ Multinomial regression med elastic net: accuracy = 88.49%
 - ▶ Support vector machine: accuracy: 89.42%

PREDIKTION I REGRESSION

- ▶ Linjär regression

$$y = \beta_0 + \beta_1 x + \varepsilon$$

- ▶ Vad blir **prediktionen** av Y för ett nytt x -värde, x_* ?

$$\mu_* = E(Y|X = x_*) = \beta_0 + \beta_1 x_*$$

som vi skattar med

$$\hat{y}_* = \hat{\beta}_0 + \hat{\beta}_1 x_*$$

- ▶ Se Baron s. 377-378 för härledning av samplingfördelning, standardfel och konfidensintervall för \hat{y}_* .
- ▶ Ett konfidensintervall för \hat{y}_* är osäkerheten om **populationens väntevärde** vid $x = x_*$. Dvs osäkerheten om regressionslinjen.
- ▶ Men hur ser osäkerheten för ett faktiskt y -värde ut om $x = x_*$?
- ▶ **Prognosintervall** för Y när $x = x_*$

$$\hat{\beta}_0 + \hat{\beta}_1 x_* \pm t_{\alpha/2} \cdot s \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_* - \bar{x})^2}{S_{xx}}}$$

BAYESIANSK PREDIKTION

- **Prediktiv fördelning** för ny observation \tilde{X} :

$$p(\tilde{x}|x_1, \dots, x_n) = \int f(\tilde{x}|x_1, \dots, x_n, \theta) \pi(\theta|x_1, \dots, x_n) d\theta$$

- Exempel: $x_1, \dots, x_n | \theta \sim N(\theta, \sigma^2)$ och σ^2 känd.
- Prediktiv fördelning

$$\tilde{x}|x_1, \dots, x_n \sim N(\mu_x, \tau_x^2 + \sigma^2)$$

- Vår bästa prognos: $\mu_x = \mathbb{E}(\theta|x_1, \dots, x_n)$.
- Prediktionsvarians = Varians pga osäkerhet om θ (τ_x^2) + Varians pga osäkerhet i populationen kring θ (σ^2).

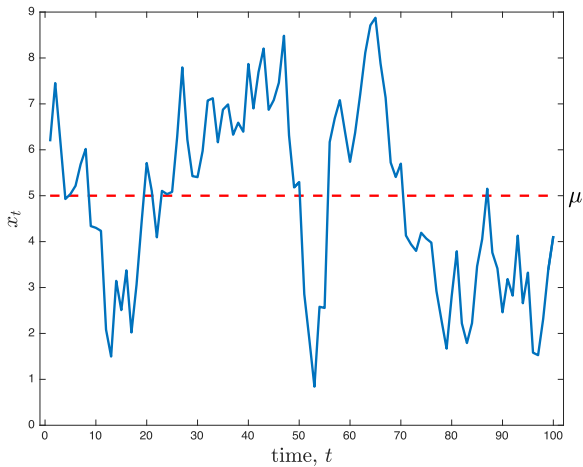
BAYESIANSK PREDIKTION - AR-PROCESS

- ▶ **Autoregressiv process** av första ordningen

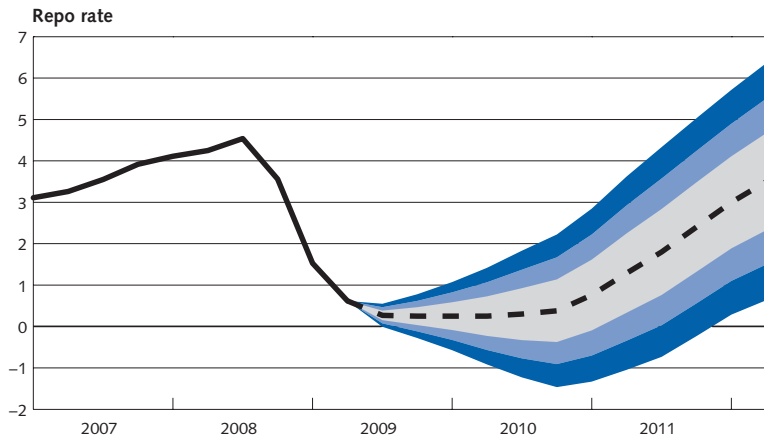
$$y_t = \mu + \phi(y_{t-1} - \mu) + \varepsilon_t, \quad \varepsilon_t \stackrel{iid}{\sim} N(0, \sigma^2)$$

- ▶ **Steady-state** (processens obetingade väntevärde): μ .
- ▶ Posterior: $\pi(\phi, \mu, \sigma | y_1, \dots, y_T)$.
- ▶ Vi är intresserade av **h -steps prognosfördelningen**:
 $p(y_{T+1}, y_{T+2}, \dots, y_{T+h} | y_1, \dots, y_T)$.
- ▶ **Simulering** från $p(y_{T+1}, y_{T+2}, \dots, y_{T+h} | y_1, \dots, y_T)$:
 1. Simulera $\theta^{(1)} = (\phi^{(1)}, \mu^{(1)}, \sigma^{(1)})$ från posteriorn $p(\phi, \mu, \sigma | y_1, \dots, y_T)$
 2. Betingat på $\theta = \theta^{(1)}$ simulera en **prognosbana**
 - ▶ $\tilde{y}_{T+1}^{(1)} \sim p(y_{T+1} | y_T, \theta^{(1)})$
 - ▶ $\tilde{y}_{T+2}^{(1)} \sim p(y_{T+2} | \tilde{y}_{T+1}^{(1)}, \theta^{(1)})$
 - ▶ \dots
 - ▶ $\tilde{y}_{T+h}^{(1)} \sim p(y_{T+h} | \tilde{y}_{T+h-1}^{(1)}, \theta^{(1)})$
- ▶ Upprepa steg 1 och 2 ett stort antal gånger.

AR(1)-PROCESS



BAYESIAN h -STEP AHEAD PREDICTIVE DISTRIBUTION



BESLUTSTEORI

- ▶ Låt $\theta \in \Theta$ vara en **okänd kvantitet, tillstånd**.
Exempel: Sjukdom, Global temperaturökning, antalet buggar.
- ▶ Låt $a \in \mathcal{A}$ vara ett **beslut** (eng. action).
Ex: operation, energisskatt, releasedatum.
- ▶ Välja beslut a när tillståndet visar sig vara θ ger dig **nyttan** (eng. utility)

$$U(a, \theta)$$

- ▶ Alternativt: **förlust** (eng. loss):

$$L(a, \theta) = -U(a, \theta)$$

DISKRETA TILLSTÅND, DISKRETA BESLUT

- Förlusttabell för problem med två möjliga θ -utfall och två möjliga beslut:

| | θ_1 | θ_2 |
|-------|--------------------|--------------------|
| a_1 | $L(a_1, \theta_1)$ | $L(a_1, \theta_2)$ |
| a_2 | $L(a_2, \theta_1)$ | $L(a_2, \theta_2)$ |

- Exempel:

| | Regnigt | Soligt |
|---------------|---------|--------|
| Paraply | 20 | 10 |
| Inget paraply | 50 | 0 |

BESLUT

- ▶ **Tillståndrummet** Θ kan vara diskret eller kontinuerligt.
- ▶ **Beslutsrummet** \mathcal{A} kan vara diskret eller kontinuerligt.
- ▶ Kontinuerligt Θ , diskret \mathcal{A} :
 - ▶ Brobygge: θ = grad av miljöpåverkan och $\mathcal{A} = \{\text{bygga, ej bygga}\}$.
- ▶ Diskret Θ , kontinuerligt \mathcal{A} :
 - ▶ θ = antalet buggar. $\theta \in \{0, 1, 2, \dots\}$ och $\mathcal{A} = \text{reasetid}$.
 - ▶ brottsdom. $\theta \in \{\text{oskyldig, skyldig}\}$. $\mathcal{A} = \{\text{tid i fängelse}\}$.
- ▶ Kontinuerligt Θ , Kontinuerligt \mathcal{A} :
 - ▶ θ = efterfrågan på produkt, $\mathcal{A} = \{\text{hur många enheter i lager?}\}$

KONTINUERLIGA TILLSTÅND, KONTINUERLIGA BESLUT

- ▶ Exempel på **förlustfunktioner** när både a och θ är kontinuerliga:

- ▶ **Linjär**: $L(a, \theta) = |a - \theta|$
- ▶ **Kvadratisk**: $L(a, \theta) = (a - \theta)^2$
- ▶ **Lin-Lin**:

$$L(a, \theta) = \begin{cases} c_1 \cdot |a - \theta| & \text{if } a \leq \theta \\ c_2 \cdot |a - \theta| & \text{if } a > \theta \end{cases}$$

- ▶ Exempel:

- ▶ θ antalet efterfrågade produkter
- ▶ a antal produkter i lager
- ▶ Nyttä

$$U(a, \theta) = \begin{cases} p \cdot \theta - c_1(a - \theta) & \text{om } a > \theta \text{ [för stort lager]} \\ p \cdot a - c_2(\theta - a)^2 & \text{om } a \leq \theta \text{ [för litet lager]} \end{cases}$$

OPTIMALA BESLUT

- ▶ Exempel på vanlig beslutsregel: **Minimax**. Välj det beslut som minimerar den maximala förlusten.
- ▶ Bayes: Välj det beslut som **maximerar förväntad nytta** a posteriori:

$$a_{bayes} = \operatorname{argmax}_{a \in \mathcal{A}} \mathbb{E}[U(a, \theta)],$$

där \mathbb{E} är väntevärdet med avseende på aposteriorn $p(\theta|Data)$.

- ▶ I praktiken: simulera $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(N)}$ från $p(\theta|Data)$ och approximera

$$\mathbb{E}[U(a, \theta)] \approx N^{-1} \sum_{i=1}^N U(a, \theta^{(i)})$$

- ▶ **Separationsprincipen:**

1. Ta först fram $p(\theta|Data)$...
2. därefter $U(a, \theta)$ och slutligen ...
3. välj det $a \in \mathcal{A}$ som maximerar $\mathbb{E}[U(a, \theta)]$.

VÄNTEVÄRDE, MEDIAN ELLER TYPVÄRDE?

- ▶ Hur kan vi bäst sammanfatta en aposteriorifördelning $p(\theta|Data)$ med ett enda tal?
- ▶ Att välja en punktskattning är ett beslutsproblem.
- ▶ Valet beror på din förlustfunktion:
 - ▶ **Linjär förlust** → Posterior median är optimal
 - ▶ **Quadratic loss** → Posteriorväntevärdet $\mathbb{E}(\theta|Data)$ är optimal
 - ▶ **Lin-Lin loss** → $c_1/(c_1 + c_2)$ kvantilen i posteriorn är optimal
 - ▶ **Noll-ett förlust** → Posterior typvärdet är optimalt

FÖRLUSTFUNKTIONER

