# Kompletne formule strojnog učenja

# 1 Osnovni koncepti

## 1.1 Primjeri, hipoteza, model

**Primjer hipoteze za regresiju**:

$$h(x; \theta_0, \theta_1) = \theta_1 x + \theta_0$$

**Primjer hipoteze za klasifikaciju pravcem**:

$$h(x_1, x_2; \theta_0, \theta_1, \theta_2) = \mathbf{1}\{\theta_1 x_1 + \theta_2 x_2 + \theta_0 \geq 0\}$$

gdje je $\mathbf{1}\{P\}$ indikatorska funkcija:

$$\mathbf{1}\{P\} = \begin{cases} 1 & \text{ako } P \equiv \top \\ 0 & \text{inače} \end{cases}$$

**Model** – skup hipoteza parametriziranih s $\theta$:

$$\mathcal{H} = \{h(x; \theta)\}_\theta$$

## 1.2 Pogreška i funkcija gubitka

**Empirijska pogreška za klasifikaciju**:

$$E(h|\mathcal{D}) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}\{h(x^{(i)}) \neq y^{(i)}\}$$

**Funkcija gubitka nula-jedan**:

$$L(y, h(x)) = \mathbf{1}\{h(x^{(i)}) \neq y^{(i)}\}$$

**Općenito, empirijska pogreška je očekivana vrijednost funkcije gubitka**:

$$E(h|\mathcal{D}) = \frac{1}{N} \sum_{i=1}^{N} L(y^{(i)}, h(x^{(i)}))$$

# 2 Regresija

## 2.1 Linearna regresija

**Jednostavna linearna regresija** ($n = 1$):

$$h(x; w_0, w_1) = w_0 + w_1 x$$

**Višestruka linearna regresija** ($n > 1$):

$$h(x; w) = w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_n x_n = \sum_{i=1}^{n} w_i x_i + w_0 = w^T x$$

**Kvadratni gubitak**:

$$L(y, h(x)) = (y - h(x))^2$$

**Funkcija pogreške za regresiju** (zbroj kvadratnih reziduala):

$$E(h|\mathcal{D}) = \frac{1}{2} \sum_{i=1}^{N} (y^{(i)} - h(x^{(i)}))^2$$

**Optimizacijski postupak**:

$$h^* = \arg\min_{h \in \mathcal{H}} E(h|\mathcal{D}) = \arg\min_{h \in \mathcal{H}} \frac{1}{2} \sum_{i=1}^{N} (y^{(i)} - h(x^{(i)}))^2$$

**Rješenje za jednostavnu regresiju** (iz $\nabla_{w_0, w_1} E = 0$):

$$w_0 = \bar{y} - w_1 \bar{x}$$

$$w_1 = \frac{\sum_{i=1}^{N} x^{(i)} y^{(i)} - N\bar{x}\bar{y}}{\sum_{i=1}^{N} (x^{(i)})^2 - N\bar{x}^2}$$

**Matrica dizajna** za višestruku regresiju:

$$X = \begin{pmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \dots & x_n^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \dots & x_n^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(N)} & x_2^{(N)} & \dots & x_n^{(N)} \end{pmatrix}$$

**Funkcija pogreške (matrični oblik)**:

$$E(w|\mathcal{D}) = \frac{1}{2}(Xw - y)^T(Xw - y) = \frac{1}{2}(w^T X^T X w - 2y^T X w + y^T y)$$

**Minimizacija**:

$$\nabla_w E = w^T X^T X - y^T X = 0$$

**Rješenje najmanjih kvadrata (OLS)**:

$$w = (X^T X)^{-1} X^T y = X^+ y$$

gdje je $X^+ = (X^T X)^{-1} X^T$ Moore-Penroseov pseudoinverz.

**Uvjet za postojanje rješenja**: $\text{rang}(X^T X) = \text{rang}(X) = n + 1$

**Gramova matrica**: $X^T X$ je $(n+1) \times (n+1)$ matrica.

## 2.2 Nelinearna regresija

**Bazne funkcije**:

$$\{\phi_0, \phi_1, \phi_2, \dots, \phi_m\}, \quad \phi_j : \mathbb{R}^n \to \mathbb{R}, \quad \phi_0(x) = 1$$

**Funkcija preslikavanja u prostor značajki**:

$$\phi : \mathbb{R}^n \to \mathbb{R}^{m+1}; \phi(x) = (\phi_0(x), \dots, \phi_m(x))$$

**Model s preslikavanjem**:

$$h(x; w) = \sum_{j=0}^{m} w_j \phi_j(x) = w^T \phi(x)$$

**Primjeri preslikavanja**:

- Linearna regresija: $\phi(x) = (1, x_1, x_2, \dots, x_n)$
- Jednostruka polinomijalna: $\phi(x) = (1, x, x^2, \dots, x^m)$
- Višestruka polinomijalna (2. stupanj): $\phi(x) = (1, x_1, x_2, x_1 x_2, x_1^2, x_2^2)$

**Matrica dizajna s preslikavanjem**:

$$\Phi = \begin{pmatrix} 1 & \phi_1(x^{(1)}) & \dots & \phi_m(x^{(1)}) \\ 1 & \phi_1(x^{(2)}) & \dots & \phi_m(x^{(2)}) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \phi_1(x^{(N)}) & \dots & \phi_m(x^{(N)}) \end{pmatrix}_{N \times (m+1)}$$

**Rješenje s preslikavanjem u prostor značajki**:

$$w = (\Phi^T \Phi)^{-1} \Phi^T y = \Phi^+ y$$

## 2.3 Regularizacija

**p-norma vektora težina**:

$$\Omega(w) = \|w\|_p = \left( \sum_{j=1}^{m} |w_j|^p \right)^{\frac{1}{p}}$$

**Različite norme**:

- $L_2$-norma ($p = 2$): $\|w\|_2 = \sqrt{\sum_{j=1}^{m} w_j^2} = \sqrt{w^T w}$
- $L_1$-norma ($p = 1$): $\|w\|_1 = \sum_{j=1}^{m} |w_j|$
- $L_0$-norma ($p = 0$): $\|w\|_0 = \sum_{j=1}^{m} \mathbf{1}\{w_j \neq 0\}$

**L2-regularizacija (Tikhonovljeva regularizacija)**:

$$E_R(w|\mathcal{D}) = \frac{1}{2}\sum_{i=1}^{N}(w^T\phi(x^{(i)}) - y^{(i)})^2 + \frac{\lambda}{2}\|w\|_2^2$$

**Hrbatna regresija (Ridge regression)** - L2 regularizacija:

$$E_R(w|\mathcal{D}) = \frac{1}{2}(\Phi w - y)^T(\Phi w - y) + \frac{\lambda}{2}w^T w$$

**Rješenje hrbatne regresije**:

$$w = (\Phi^T\Phi + \lambda I)^{-1}\Phi^T y$$

**L1-regularizacija (LASSO)**:

$$E_R(w|\mathcal{D}) = \frac{1}{2}\sum_{i=1}^{N}(w^T\phi(x^{(i)}) - y^{(i)})^2 + \frac{\lambda}{2}\|w\|_1$$

**L0-regularizacija**:

$$E_R(w|\mathcal{D}) = \frac{1}{2}\sum_{i=1}^{N}(w^T\phi(x^{(i)}) - y^{(i)})^2 + \frac{\lambda}{2}\sum_{j=1}^{m}\mathbf{1}\{w_j \neq 0\}$$

## 2.4 Probabilistička interpretacija

**Model sa šumom**:

$$y^{(i)} = f(x^{(i)}) + \epsilon_i, \quad \text{gdje je } \epsilon \sim \mathcal{N}(0, \sigma^2)$$

**Normalna razdioba**:

$$p(Y = y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$$

**Vjerojatnost oznake za zadani primjer**:

$$p(y|x) = \mathcal{N}(f(x), \sigma^2)$$

**Vjerojatnost cijelog skupa primjera** (uz iid pretpostavku):

$$p(y|X) = \prod_{i=1}^{N} p(y^{(i)}|x^{(i)})$$

**Log-izglednost težina** $w$:

$$\ln p(y|X, w) = \ln\prod_{i=1}^{N}\mathcal{N}(h(x^{(i)}; w), \sigma^2) \propto -\frac{1}{2}\sum_{i=1}^{N}(y^{(i)} - h(x^{(i)}; w))^2$$

# 3 Linearni diskriminativni modeli

**Linearni diskriminativni model**:

$$h(x; w) = w^T x$$

**Granica između klasa**: $h(x) = 0$ (ponekad $h(x) = 0.5$).
**Geometrija linearnog modela** (za $n = 2$):

$$w_1x_1 + w_2x_2 + w_0 = 0$$

**Normala hiperravnine**: $w$ je normala hiperravnine.
**Udaljenost primjera od hiperravnine**:

$$d = \frac{h(x)}{\|w\|} = \frac{w^T x + w_0}{\|w\|}$$

**Višeklasna klasifikacija**:

- Jedan-naspram-jedan (OVO): $\binom{K}{2}$ binarnih modela
- Jedan-naspram-ostali (OVR): $K$ binarnih modela

**OVO shema**:

$$h(x) = \arg\max_i \sum_{j \neq i} \mathrm{sgn}(h_{ij}(x))$$

**OVR shema**:

$$h(x) = \arg\max_j h_j(x)$$

### 3.1 Klasifikacija regresijom

**Funkcija pogreške**:

$$E(w|\mathcal{D}) = \frac{1}{2} \sum_{i=1}^{N} (w^T \phi(x^{(i)}) - y^{(i)})^2 = \frac{1}{2} (\Phi w - y)^T (\Phi w - y)$$

**Model za klasifikaciju regresijom**:

$$h(x; w) = \mathbf{1}\{w^T \phi(x) \geq 0.5\}$$

### 3.2 Perceptron

**Perceptron model**:

$$h(x) = f(w^T \phi(x))$$

**Funkcija praga kao aktivacijska funkcija**:

$$f(\alpha) = \begin{cases} +1 & \text{ako } \alpha \geq 0 \\ -1 & \text{inače} \end{cases}$$

**Gubitak perceptrona**:

$$L(y, h(x)) = \max(0, -w^T \phi(x) y)$$

**Funkcija pogreške perceptrona**:

$$E(w|\mathcal{D}) = - \sum_{i: f(w^T \phi(x^{(i)})) \neq y^{(i)}} w^T \phi(x^{(i)}) y^{(i)} = \sum_{i=1}^{N} \max(0, -w^T \phi(x^{(i)}) y^{(i)})$$

**Gradijent funkcije gubitka**:

$$\nabla_w L = -\phi(x) y$$

**Widrow-Hoffovo pravilo**:

$$w \leftarrow w + \eta \phi(x) y$$

## 4 Logistička regresija

**Logistička (sigmoidalna) funkcija**:

$$\sigma(\alpha) = \frac{1}{1 + \exp(-\alpha)}$$

**Derivacija sigmoide**:

$$\frac{\partial \sigma(\alpha)}{\partial \alpha} = \sigma(\alpha)(1 - \sigma(\alpha))$$

**Model logističke regresije**:

$$h(x; w) = \sigma(w^T \phi(x)) = \frac{1}{1 + \exp(-w^T \phi(x))} = P(y = 1|x)$$

**Bernoullijeva razdioba izlaza**:

$$P(y|\mu) = \begin{cases} \mu & \text{ako } y = 1 \\ 1 - \mu & \text{inače} \end{cases} = \mu^y (1 - \mu)^{1-y}$$

**Vjerojatnost primjera**:

$$P(y^{(i)}|x^{(i)}) = h(x^{(i)}; w)^{y^{(i)}}(1 - h(x^{(i)}; w))^{1-y^{(i)}}$$

**Log-izglednost oznaka**:

$$\ln P(y|X, w) = \sum_{i=1}^{N} \Big( y^{(i)} \ln h(x^{(i)}; w) + (1 - y^{(i)}) \ln(1 - h(x^{(i)}; w)) \Big)$$

**Empirijska pogreška** (negativna log-izglednost):

$$E(w|\mathcal{D}) = \frac{1}{N} \sum_{i=1}^{N} \Big( -y^{(i)} \ln h(x^{(i)}; w) - (1 - y^{(i)}) \ln(1 - h(x^{(i)}; w)) \Big)$$

**Gubitak unakrsne entropije**:

$$L(y, h(x)) = -y \ln h(x) - (1 - y) \ln(1 - h(x))$$

**Logistički gubitak** (za $y \in \{-1, +1\}$):

$$L(y, h(x)) = \frac{1}{\ln 2} \ln(1 + \exp(-y w^T \phi(x)))$$

**Gradijent funkcije gubitka**:

$$\nabla_w L(y, h(x)) = (h(x) - y)\phi(x)$$

**Gradijent funkcije pogreške**:

$$\nabla_w E(w|\mathcal{D}) = \sum_{i=1}^{N} (h(x^{(i)}) - y^{(i)})\phi(x^{(i)})$$

## 4.1 Gradijentni spust

**Općenito pravilo gradijentnog spusta**:

$$x \leftarrow x - \eta \nabla f(x)$$

**Grupni gradijentni spust**:

$$w \leftarrow w - \eta \sum_{i=1}^{N} \nabla L(y^{(i)}, h(x^{(i)}; w))$$

**Stohastički gradijentni spust (SGD)**:

$$w \leftarrow w - \eta \nabla L(y^{(i)}, h(x^{(i)}; w))$$

## 4.2 Regularizirana logistička regresija

**L2-regularizirana pogreška**:

$$E_R(w|\mathcal{D}) = \sum_{i=1}^{N} \Big( -y^{(i)} \ln h(x^{(i)}) - (1 - y^{(i)}) \ln(1 - h(x^{(i)})) \Big) + \frac{\lambda}{2} w^T w$$

**Ažuriranje težina za L2 regularizaciju**:

$$w \leftarrow w(1 - \eta\lambda) - \eta \sum_{i=1}^{N} (h(x^{(i)}) - y^{(i)})\phi(x^{(i)})$$

## 4.3 Višeklasna logistička regresija

**Softmax funkcija**:

$$\text{softmax}_k(x_1, \ldots, x_K) = \frac{\exp(x_k)}{\sum_{j=1}^{K} \exp(x_j)}$$

**Multinomijalna logistička regresija**:

$$h_k(x; W) = \frac{\exp(w_k^T \phi(x))}{\sum_{j=1}^{K} \exp(w_j^T \phi(x))} = P(y = k|x, W)$$

**Multinomijalna razdioba**:

$$P(y|\mu) = \prod_{k=1}^{K} \mu_k^{y_k}$$

**Poopćena pogreška unakrsne entropije**:

$$E(W|\mathcal{D}) = -\sum_{i=1}^{N}\sum_{k=1}^{K} y_k^{(i)} \ln h_k(x^{(i)}; W)$$

**Gradijent za klasu** $k$:

$$\nabla_{w_k} E(W|\mathcal{D}) = \sum_{i=1}^{N} (h_k(x^{(i)}; W) - y_k^{(i)})\phi(x^{(i)})$$

# 5 Stroj potpornih vektora (SVM)

**SVM linearan model**:

$$h(x; w, w_0) = w^T x + w_0$$

**Linearna odvojivost** (uz $y \in \{-1, +1\}$):

$$\forall (x^{(i)}, y^{(i)}) \in \mathcal{D} : y^{(i)} h(x^{(i)}) \geq 0$$

**Udaljenost primjera od hiperravnine**:

$$d^{(i)} = \frac{1}{\|w\|} y^{(i)} (w^T x^{(i)} + w_0)$$

**Problem maksimalne margine**:

$$\underset{w, w_0}{\arg\max} \left\{ \frac{1}{\|w\|} \min_i y^{(i)} (w^T x^{(i)} + w_0) \right\}$$

**Normalizacija** (za primjere najbliže margini):

$$y^{(i)} (w^T x^{(i)} + w_0) = 1$$

**SVM primarni problem** (tvrda margina):

$$\underset{w, w_0}{\arg\min} \frac{1}{2} \|w\|^2 \quad \text{uz } y^{(i)} (w^T x^{(i)} + w_0) \geq 1, \quad \forall i$$

## 5.1 Lagrangeova dualnost

**Lagrangeova funkcija**:

$$\mathcal{L}(w, w_0, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^{N} \alpha_i \left( y^{(i)} (w^T x^{(i)} + w_0) - 1 \right)$$

**Uvjeti iz derivacija**:

$$\frac{\partial \mathcal{L}}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{N} \alpha_i y^{(i)} x^{(i)}$$

$$\frac{\partial \mathcal{L}}{\partial w_0} = 0 \Rightarrow \sum_{i=1}^{N} \alpha_i y^{(i)} = 0$$

**Dualna Lagrangeova funkcija**:

$$\tilde{\mathcal{L}}(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y^{(i)} y^{(j)} (x^{(i)})^T x^{(j)}$$

**SVM dualni problem**:

$$\underset{\alpha}{\arg\max} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y^{(i)} y^{(j)} (x^{(i)})^T x^{(j)}$$

$$\text{uz } \alpha_i \geq 0 \text{ i } \sum_{i=1}^{N} \alpha_i y^{(i)} = 0$$

**KKT uvjeti**:

$$y^{(i)}(w^T x^{(i)} + w_0) \geq 1, \quad \alpha_i \geq 0, \quad \alpha_i(y^{(i)} h(x^{(i)}) - 1) = 0$$

**Dualni SVM model** (za predikciju):

$$h(x) = \sum_{i=1}^{N} \alpha_i y^{(i)} (x^{(i)})^T x + w_0$$

### 5.2 Meka margina

**Reformulacija ograničenja** (dopuštanje ulaska u marginu):

$$y^{(i)}(w^T x^{(i)} + w_0) \geq 1 - \xi_i, \quad \xi_i \geq 0$$

**SVM meke margine (primarni problem)**:

$$\underset{w, w_0, \xi}{\arg\min} \left\{ \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} \xi_i \right\}$$

$$\text{uz } y^{(i)}(w^T x^{(i)} + w_0) \geq 1 - \xi_i \text{ i } \xi_i \geq 0, \quad \forall i$$

**Lagrangeova funkcija za meku marginu**:

$$\mathcal{L}(w, w_0, \xi, \alpha, \beta) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \alpha_i (y^{(i)}(w^T x^{(i)} + w_0) - 1 + \xi_i) - \sum_{i=1}^{N} \beta_i \xi_i$$

**Dodatni uvjet iz derivacija**:

$$\frac{\partial \mathcal{L}}{\partial \xi_i} = 0 \Rightarrow \alpha_i = C - \beta_i$$

**Dualni problem za meku marginu**: Isti kao za tvrdu marginu, uz dodatno ograničenje $0 \leq \alpha_i \leq C$.
**Gubitak zglobnice (hinge loss)**:

$$L(y, h(x)) = \max(0, 1 - y h(x))$$

**Alternativna formulacija SVM ciljne funkcije**:

$$E(w|\mathcal{D}) = \sum_{i=1}^{N} \max(0, 1 - y^{(i)} h(x^{(i)})) + \frac{\lambda}{2} \|w\|^2 \quad (\text{gdje je } \lambda = 1/C)$$

## 6 Jezgrene metode

**Jezgrena funkcija** (mjera sličnosti):

$$\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$$

**Jezgreni trik**:

$$\kappa(x, x') = \phi(x)^T \phi(x')$$

**Primjeri jezgri**:

- Linearna jezgra: $\kappa(x, x') = x^T x'$
- Gaussova RBF-jezgra: $\kappa(x, x') = \exp(-\gamma \|x - x'\|^2)$
- Eksponencijalna jezgra: $\kappa(x, x') = \exp(-\gamma \|x - x'\|)$
- Inverzna kvadratna jezgra: $\kappa(x, x') = \frac{1}{1 + \|x - x'\|^2}$
- Polinomijalna jezgra: $\kappa(x, x') = (\gamma x^T x' + c)^d$
- RBF s Mahalanobisovom udaljenosti: $\kappa(x, x') = \exp(-\frac{1}{2}(x - x')^T \Sigma^{-1} (x - x'))$

**Jezgreni stroj** (model s jezgrama kao baznim funkcijama):

$$\phi(x) = (1, \kappa(x, \mu_1), \kappa(x, \mu_2), \ldots, \kappa(x, \mu_m))$$

**Jezgreni SVM model** (za predikciju):

$$h(x) = \sum_{i=1}^{N} \alpha_i y^{(i)} \kappa(x, x^{(i)}) + w_0$$

**Jezgreni SVM dualni problem**:

$$\operatorname*{arg\,max}_{\alpha} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y^{(i)} y^{(j)} \kappa(x^{(i)}, x^{(j)})$$

**Jezgrena (Gramova) matrica**:

$$K_{ij} = \kappa(x^{(i)}, x^{(j)}) \Rightarrow K = \Phi\Phi^T$$

# 7 Neparametarske metode

**k-NN klasifikator**:

$$h(x) = \operatorname*{arg\,max}_{j \in \{0,\ldots,K-1\}} \sum_{(x^{(i)}, y^{(i)}) \in \mathrm{NN}_k(x)} \mathbf{1}\{y^{(i)} = j\}$$

**Težinski k-NN klasifikator**:

$$h(x) = \operatorname*{arg\,max}_{j \in \{0,\ldots,K-1\}} \sum_{i=1}^{N} \kappa(x^{(i)}, x) \mathbf{1}\{y^{(i)} = j\}$$

**k-NN zaglađivač (regresija)**:

$$h(x) = \frac{1}{k} \sum_{(x^{(i)}, y^{(i)}) \in \mathrm{NN}_k(x)} y^{(i)}$$

**Jezgreni zaglađivač (regresija)**:

$$h(x) = \frac{\sum_{i=1}^{N} \kappa(x^{(i)}, x) y^{(i)}}{\sum_{i=1}^{N} \kappa(x^{(i)}, x)}$$

# 8 Osnove vjerojatnosti i procjene parametara

## 8.1 Slučajne varijable

**Očekivanje**:

$$E[X] = \sum_x x P(x) \quad \text{ili} \quad E[X] = \int_{-\infty}^{\infty} x p(x) dx$$

**Varijanca**:

$$\mathrm{Var}(X) = \sigma_X^2 = E[(X - E[X])^2]$$

**Kovarijanca**:

$$\mathrm{Cov}(X, Y) = \sigma_{X,Y} = E[(X - E[X])(Y - E[Y])]$$

**Pearsonov koeficijent korelacije**:

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y} \in [-1, +1]$$

**Matrica kovarijacije**:

$$\Sigma = E[(X - E[X])(X - E[X])^T] = \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \ldots \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) & \ldots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

## 8.2 Distribucije

**Bernoullijeva razdioba**:

$$P(x|\mu) = \mu^x(1-\mu)^{1-x}$$

**Kategorička (Multinulijeva) razdioba**:

$$P(x|\mu) = \prod_{k=1}^{K} \mu_k^{x_k}, \quad \text{gdje je } x \text{ one-hot vektor}$$

**Normalna (Gaussova) razdioba**:

$$p(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

**Multivarijatna normalna razdioba**:

$$p(x|\mu, \Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

**Mahalanobisova udaljenost (kvadrat)**:

$$\Delta^2 = (x-\mu)^T\Sigma^{-1}(x-\mu)$$

**Beta-distribucija**:

$$p(\mu|\alpha, \beta) = \frac{1}{B(\alpha, \beta)}\mu^{\alpha-1}(1-\mu)^{\beta-1}$$

**Dirichletova distribucija**:

$$P(\mu|\alpha) = \frac{1}{B(\alpha)} \prod_{k=1}^{K} \mu_k^{\alpha_k - 1}, \quad \text{uz } \sum \mu_k = 1$$

## 8.3 Procjenitelji

**Pristranost (bias) procjenitelja $\Theta$**:

$$b(\Theta) = E[\Theta] - \theta$$

**Funkcija izglednosti (Likelihood)**:

$$L(\theta|\mathcal{D}) = p(\mathcal{D}|\theta) = \prod_{i=1}^{N} p(x^{(i)}|\theta)$$

**MLE - Procjenitelj najveće izglednosti**:

$$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} L(\theta|\mathcal{D}) = \arg\max_{\theta} \ln L(\theta|\mathcal{D})$$

**MLE za parametre normalne razdiobe**:

$$\hat{\mu}_{\text{MLE}} = \frac{1}{N}\sum_{i=1}^{N} x^{(i)}, \quad \hat{\sigma}^2_{\text{MLE}} = \frac{1}{N}\sum_{i=1}^{N}(x^{(i)} - \hat{\mu}_{\text{MLE}})^2$$

**Aposteriorna vjerojatnost parametra**:

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})} \propto L(\theta|\mathcal{D})p(\theta)$$

**MAP - Maximum A Posteriori procjenitelj**:

$$\hat{\theta}_{\text{MAP}} = \arg\max_{\theta} p(\theta|\mathcal{D}) = \arg\max_{\theta} \ln(L(\theta|\mathcal{D})p(\theta))$$

**MAP za Bernoullijevu (Beta apriorna)**:

$$\hat{\mu}_{\text{MAP}} = \frac{m + \alpha - 1}{N + \alpha + \beta - 2}$$

**Laplaceovo zaglađivanje** (MAP za Bern. uz $\alpha = \beta = 2$):

$$\hat{\mu}_{\text{MAP}} = \frac{m+1}{N+2}$$

**MAP za kategoričku** (Dirichlet apriorna, $\alpha_k = 2$ za sve $k$):

$$\hat{\mu}_{k,\text{MAP}} = \frac{N_k + 1}{N + K}$$

# 9 Bayesov klasifikator i PGM-ovi

## 9.1 Pravila vjerojatnosti

**Pravilo zbroja**:

$$P(x) = \sum_y P(x, y)$$

**Pravilo umnoška**:

$$P(x, y) = P(y|x)P(x) = P(x|y)P(y)$$

**Bayesovo pravilo**:

$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

**Pravilo lanca**:

$$P(x_1, \dots, x_n) = P(x_1)P(x_2|x_1) \dots P(x_n|x_1, \dots, x_{n-1}) = \prod_{k=1}^{n} P(x_k|x_1, \dots, x_{k-1})$$

## 9.2 Bayesov klasifikator

**Model Bayesovog klasifikatora**:

$$h_j(x; \theta) = P(y = j|x) = \frac{p(x|y = j)P(y = j)}{\sum_k p(x|y = k)P(y = k)}$$

**MAP-hipoteza**:

$$h(x) = \arg\max_j p(x|y = j)P(y = j)$$

**Log-vjerojatnost za Gaussov klasifikator**:

$$h_j(x) = \ln p(x|y = j) + \ln P(y = j) \propto -\frac{1}{2} \ln |\Sigma_j| - \frac{1}{2}(x - \mu_j)^T \Sigma_j^{-1}(x - \mu_j) + \ln P(y = j)$$

**Pretpostavka uvjetne nezavisnosti (Naivni Bayes)**:

$$p(x_1, \dots, x_n|y) = \prod_{k=1}^{n} p(x_k|y)$$

**Naivni Bayesov klasifikator**:

$$h(x_1, \dots, x_n) = \arg\max_j P(y = j) \prod_{k=1}^{n} P(x_k|y = j)$$

## 9.3 Probabilistički grafički modeli (PGM)

**Faktorizacija Bayesove mreže**:

$$p(x) = \prod_{k=1}^{n} p(x_k|\mathrm{pa}(x_k))$$

**Uvjetna nezavisnost $X \perp Y|Z$**:

$$P(X, Y|Z) = P(X|Z)P(Y|Z) \Leftrightarrow P(X|Y, Z) = P(X|Z)$$

**Zaključivanje (posteriorni upit)**:

$$p(x_q|x_o) = \frac{\sum_{x_n} p(x_q, x_o, x_n)}{p(x_o)}$$

**Zaključivanje (MAP-upit)**:

$$x_q^* = \arg\max_{x_q} \sum_{x_n} p(x_q, x_o, x_n)$$

# 10 Grupiranje

**K-sredina ciljna funkcija**:

$$J = \sum_{k=1}^{K} \sum_{i=1}^{N} b_k^{(i)} \|x^{(i)} - \mu_k\|^2$$

**Ažuriranje centroida K-sredina**:

$$\mu_k = \frac{\sum_{i=1}^{N} b_k^{(i)} x^{(i)}}{\sum_{i=1}^{N} b_k^{(i)}}$$

**Model Gaussove mješavine (GMM)**:

$$p(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k)$$

**GMM odgovornost** (E-korak):

$$h_k^{(i)} = P(y = k|x^{(i)}) = \frac{\pi_k p(x^{(i)}|\theta_k)}{\sum_{j=1}^{K} \pi_j p(x^{(i)}|\theta_j)}$$

**GMM ažuriranje parametara** (M-korak):

$$\begin{aligned}
\mu_k &\leftarrow \frac{\sum_i h_k^{(i)} x^{(i)}}{\sum_i h_k^{(i)}} \\
\Sigma_k &\leftarrow \frac{\sum_i h_k^{(i)} (x^{(i)} - \mu_k)(x^{(i)} - \mu_k)^T}{\sum_i h_k^{(i)}} \\
\pi_k &\leftarrow \frac{1}{N} \sum_i h_k^{(i)}
\end{aligned}$$

**Udaljenosti za HAC**:

- Jednostruko povezivanje: $d_{\min}(G_i, G_j) = \min_{x \in G_i, x' \in G_j} d(x, x')$
- Potpuno povezivanje: $d_{\max}(G_i, G_j) = \max_{x \in G_i, x' \in G_j} d(x, x')$
- Prosječno povezivanje: $d_{\text{avg}}(G_i, G_j) = \frac{1}{N_i N_j} \sum_{x \in G_i} \sum_{x' \in G_j} d(x, x')$

**Silueta** (mjera kvalitete grupiranja):

$$s^{(i)} = \frac{b^{(i)} - a^{(i)}}{\max(a^{(i)}, b^{(i)})}$$

**Randov indeks**:

$$R = \frac{a + b}{\binom{N}{2}}$$

# 11 Vrednovanje modela

**Matrica zabune**:

$$\begin{pmatrix} \text{TP} & \text{FP} \\ \text{FN} & \text{TN} \end{pmatrix}$$

**Točnost (Accuracy)**:

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

**Preciznost (Precision)**:

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

**Odziv (Recall, TPR)**:

$$R = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

**Stopa lažno pozitivnih (FPR)**:

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

**Specifičnost**:

$$S = \frac{\text{TN}}{\text{TN} + \text{FP}} = 1 - \text{FPR}$$

**F1 mjera**:

$$F_1 = \frac{2PR}{P + R}$$

**F-beta mjera**:

$$F_\beta = \frac{(1 + \beta^2)PR}{\beta^2 P + R}$$

**z-statistika**:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{N}} \sim \mathcal{N}(0, 1)$$

**t-statistika**:

$$t = \frac{\bar{x} - \mu}{\hat{\sigma}/\sqrt{N}} \sim t(N - 1)$$

**Interval pouzdanosti** (uz t-distribuciju):

$$\mu = \bar{x} \pm t_{\alpha/2, N-1} \frac{\hat{\sigma}}{\sqrt{N}}$$

**Upareni t-test**:

$$t = \frac{\bar{d}}{\hat{\sigma}_d/\sqrt{K}}$$

gdje je $\bar{d}$ prosječna razlika performansi na $K$ preklopa.

# 12 Odabir značajki

**Uzajamna informacija** (mjera zavisnosti):

$$I(x, y) = \sum_{x,y} P(x, y) \ln \frac{P(x, y)}{P(x)P(y)} = D_{\text{KL}}(P(x, y) \| P(x)P(y))$$

**t-statistika** (za usporedbu srednjih vrijednosti značajke $x_k$):

$$t = \frac{\bar{x}_{0k} - \bar{x}_{1k}}{\hat{\sigma}_i \sqrt{\frac{1}{N_0} + \frac{1}{N_1}}}$$

$\chi^2$**-statistika** (za kategoričke značajke):

$$\chi^2 = \sum_{i=1}^{K_k} \sum_{j=1}^{K} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}$$

**Ažuriranje težina za RELIEF algoritam**:

$$w_k \leftarrow w_k - \frac{1}{N}(x_k - x_k^h)^2 + \frac{1}{N}(x_k - x_k^m)^2$$

**Faktor inflacije varijance (VIF)**:

$$\text{VIF}_k = \frac{1}{1 - R_k^2}$$

**CFS vrijednost podskupa značajki**:

$$\text{Merit}_S = \frac{k\bar{r}_{x,y}}{\sqrt{k + k(k-1)\bar{r}_{x,x}}}$$