Classification

Probabilistic Generative Models I

T.L. Grobler
(with some edits by S. Kroon)

Classification

- Given $\mathbf{x}$ assign to one of k classes:
  - C_j, $j = 1,\ldots,k$
  - Assign prob $P(C_j|\mathbf{x})$
  - $C^* = \text{argmax}_{C_j} P(C_j|\mathbf{x})$
- Class prob, more useful than knowing max class prob.

Data Description

- D: $(\mathbf{x}_j, y_j)$, $j = 1,\ldots,N$
  - observation $\mathbf{x}_j$ comes with class label y_j
  - $y_j = C_j$ if $\mathbf{x}_j \in C_j$
- Constructing $P(C_j|\mathbf{x})$ given D
- Two approaches: *generative* and *discriminative*

Generative Approach

- Estimate class-conditionals: $p(\mathbf{x}|C_j)$
- Posterior (Bayes Theorem):
  - $P(C_j|\mathbf{x}) \propto p(\mathbf{x}|C_j)P(C_j)$
  - Introduce $P(C_j)$
- Probabilistic Generative Models (PGM)

Key steps

- Step 1: **Expand posterior** using logistic/softmax functions.
- Step 2: Investigate/**determine form** of arguments to logistic/softmax functions (under model assumptions).
- Step 3: **estimate parameters** for the resulting form.

Two Classes

$$
\begin{aligned}
P(C_1|\mathbf{x}) &= \frac{p(\mathbf{x}|C_1)P(C_1)}{p(\mathbf{x}|C_1)P(C_1) + p(\mathbf{x}|C_2)P(C_2)} \\
&= \frac{1}{1 + \frac{p(\mathbf{x}|C_2)P(C_2)}{p(\mathbf{x}|C_1)P(C_1)}} \\
&= \frac{1}{1 + \exp(-a(\mathbf{x}))} \\
&= \sigma(a(\mathbf{x}))
\end{aligned}
$$

$$P(C_1|\mathbf{x}) + P(C_2|\mathbf{x}) = 1$$

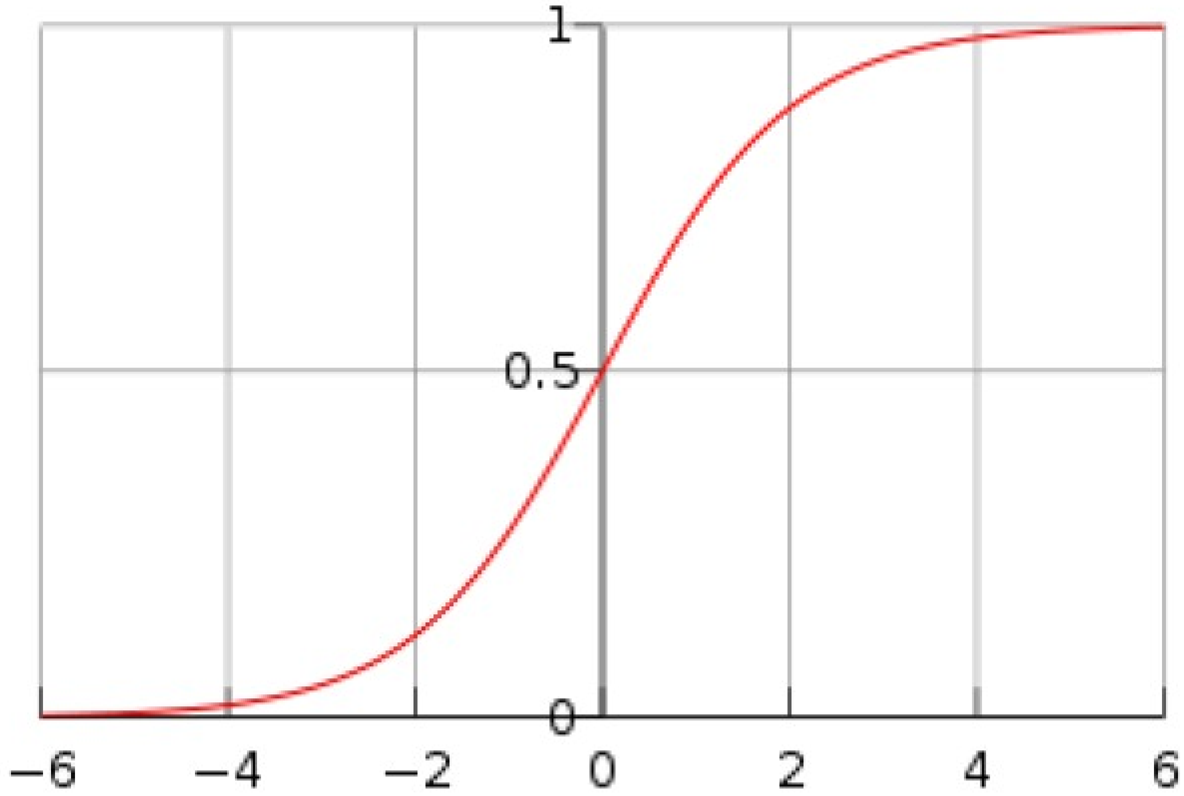
Log Posterior Odds

$$\text{odds} = \frac{p(\mathbf{x}|C_1)P(C_1)}{p(\mathbf{x}|C_2)P(C_2)}$$

$$a(\mathbf{x}) = \ln \frac{p(\mathbf{x}|\mathcal{C}_1)P(\mathcal{C}_1)}{p(\mathbf{x}|\mathcal{C}_2)P(\mathcal{C}_2)}$$

$$\boxed{\exp\left(-\ln\left(\frac{k_1}{k_2}\right)\right) = \exp\left(\ln\left(\frac{k_2}{k_1}\right)\right) = \frac{k_2}{k_1}}$$

Logistic Sigmoid Function



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

Discussion

- Logistic function:
  - a(**x**) –> [0,1]
  - Assigns posterior prob to **x**
- Odds:
  - Classification
  - C_1 if odds > 1 otherwise C_2

k Classes

$$P(C_n|\mathbf{x}) = \frac{p(\mathbf{x}|C_n)P(C_n)}{\sum_{j=1}^{k} p(\mathbf{x}|C_j)P(C_j)}$$

$$= \frac{\exp(a_n(\mathbf{x}))}{\sum_{j=1}^{k} \exp(a_j(\mathbf{x}))}$$

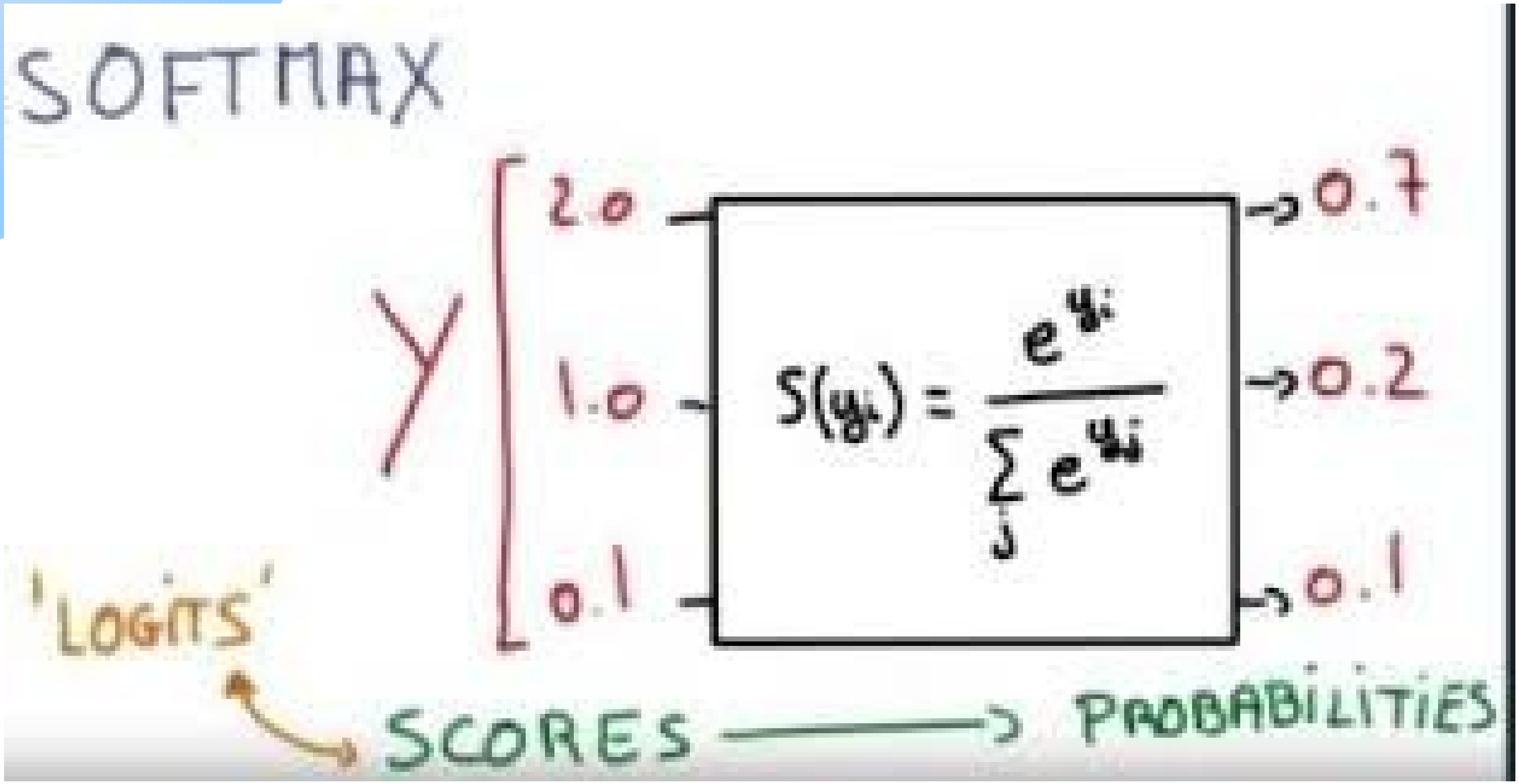
softmax

$$a_j(\mathbf{x}) = \ln p(\mathbf{x}|C_j) + \ln P(C_j)$$

$$\sum_{j=1}^{k} P(C_j|\mathbf{x}) = 1$$

$$\ln(AB) = \ln(A) + \ln(B)$$

softmax



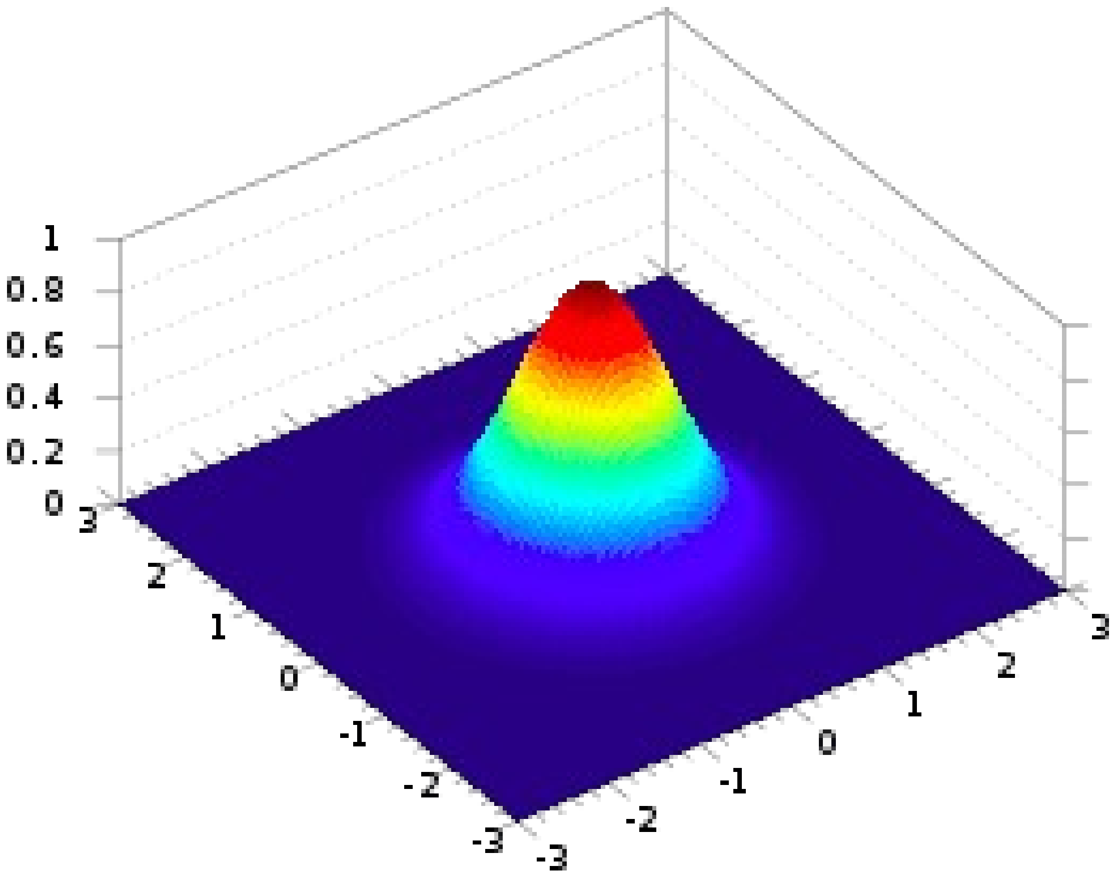
- Outputs posterior probabilities
- Emphasizes max values
- Differentiable or smooth

Note

- PGM requires:
  - Prior class probabilities $P(C_j)$
  - Class-conditional densities $p(\mathbf{x}|C_j)$

Gaussian class-conditional pdfs

$$p(\mathbf{x}|C_j) = \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{u}_j)^T \Sigma^{-1} (\mathbf{x}-\mathbf{u}_j) \right)$$



Two Classes: Shared Σ

$$P(C_1|\mathbf{x}) = \sigma(\mathbf{w}^T\mathbf{x} + w_0)$$

$$a(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0$$

$$\mathbf{w} = \Sigma^{-1}(\mathbf{u}_1 - \mathbf{u}_2)$$

$$w_0 = -\frac{1}{2}\mathbf{u}_1^T\Sigma^{-1}\mathbf{u}_1 + \frac{1}{2}\mathbf{u}_2^T\Sigma^{-1}\mathbf{u}_2 + \ln\frac{P(C_1)}{P(C_2)}$$

$$\boxed{(\mathbf{x}-\mathbf{u})^T\mathbf{A}(\mathbf{x}-\mathbf{u}) = \mathbf{x}^T\mathbf{A}\mathbf{x} - 2\mathbf{u}^T\mathbf{A}\mathbf{x} + \mathbf{u}^T\mathbf{A}\mathbf{u}}$$

Log identities

$$a(\mathbf{x}) = \ln \frac{P(\mathbf{x}|C_1)P(C_1)}{P(\mathbf{x}|C_2)P(C_2)}$$

$$= \ln P(\mathbf{x}|C_1) - \ln P(\mathbf{x}|C_2) + \ln \frac{P(C_1)}{P(C_2)}$$

Symmetric: Σ

$$= -\frac{1}{2}(\mathbf{x}-\mathbf{u}_1)^T\Sigma^{-1}(\mathbf{x}-\mathbf{u}_1) +$$

$$\frac{1}{2}(\mathbf{x}-\mathbf{u}_2)^T\Sigma^{-1}(\mathbf{x}-\mathbf{u}_2) + \ln \frac{P(C_1)}{P(C_2)}$$

$$= -\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x} + \mathbf{u}_1^T\Sigma^{-1}\mathbf{x} - \frac{1}{2}\mathbf{u}_1^T\Sigma^{-1}\mathbf{u}_1$$

$$\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x} - \mathbf{u}_2^T\Sigma^{-1}\mathbf{x} + \frac{1}{2}\mathbf{u}_2^T\Sigma^{-1}\mathbf{u}_2 + \ln \frac{P(C_1)}{P(C_2)}$$

$$= \mathbf{w}^T\mathbf{x} + w_0$$

Discussion

- Prior –> w_0
- Shared Σ
  - Quadratic Terms Cancel
  - Linear Classifier
- Parameter reduction:
  - d dim
  - $2d+1/2d(d+1)$: $\mathbf{u}_1 \mathbf{u}_2\ \Sigma$ - *generative*
  - $d+1$: $\mathbf{w}, w_0$ - *discriminative*

Linear Decision Boundary

$$
\begin{aligned}
P(C_1|\mathbf{x}) &= P(C_2|\mathbf{x}) = 1 - P(C_1|\mathbf{x}) \\
\sigma(\mathbf{w}^T\mathbf{x} + w_0) &= 1 - \sigma(\mathbf{w}^T\mathbf{x} + w_0) \\
&= \frac{1}{2} \\
\mathbf{w}^T\mathbf{x} + w_0 &= 0
\end{aligned}
$$

k-Classes: Shared Σ

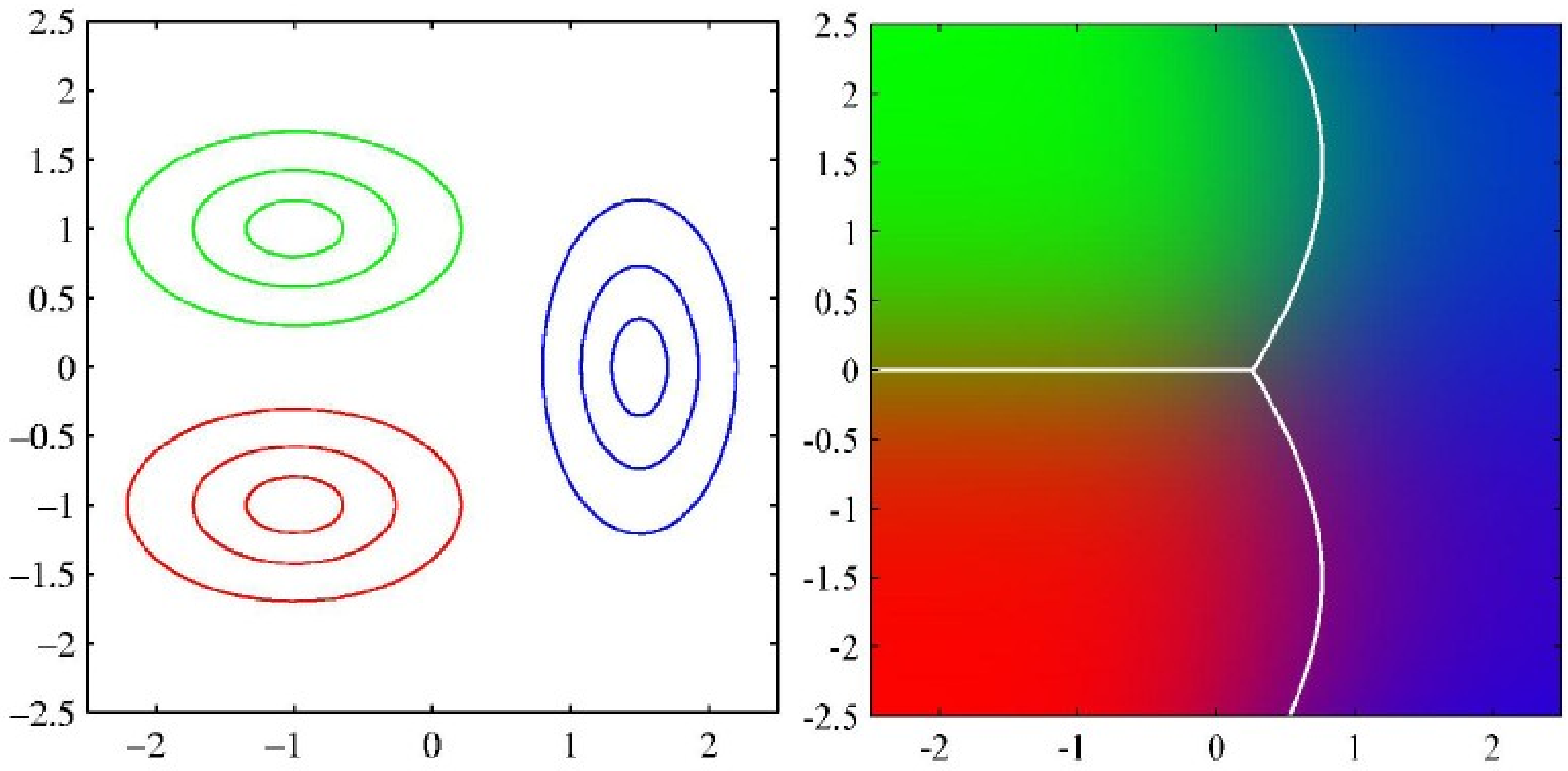
$$P(C_n|\mathbf{x}) = \frac{\exp a_n(\mathbf{x})}{\sum_{j=1}^{k} \exp a_j(\mathbf{x})}$$

$$a_j(\mathbf{x}) = \mathbf{w}_j^T \mathbf{x} + w_{j0}$$

$$\mathbf{w}_j = \Sigma^{-1}\mathbf{u}_j$$

$$w_{j0} = -\frac{1}{2}\mathbf{u}_j^T \Sigma^{-1}\mathbf{u}_j + \ln P(C_j)$$

Non-shared Σ



Non-linear Classifier: Quadratic Decision Boundary