# Foundations of Machine Learning
# AI2000 and AI5000

FoML-14
Probabilistic Generative Models - Discrete features

Dr. Konda Reddy Mopuri
Department of AI, IIT Hyderabad
July-Nov 2025

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Data-driven Intelligence
& Learning Lab

# So far in FoML

- What is ML and the learning paradigms
- Probability refresher
- MLE, MAP, and fully Bayesian treatment
- Linear Regression with basis functions - regularization & model selection
- Bias-Variance Decomposition/Tradeoff (Bayesian Regression)
- Decision Theory - three broad classification strategies
- Probabilistic Generative Models - Continuous data

# Probabilistic Generative Models - Discrete features

# Probabilistic Generative Models - Discrete

- Input: discrete feature vectors $\mathbf{x}_n = (x_1, \ldots, x_D)^T$

$x_i \in \{v_1, v_2 \ldots v_M\}$ $\underline{x} \in \{v_1, v_2 \ldots v_M\}^D$

- For simplicity, consider binary feature values $x_i \in \{0, 1\}$

$\underline{x} \in \{0, 1\}^D$

# Probabilistic Generative Models - Discrete

- For D-dim input
  - The no. of parameters to express each class conditional density $p(\mathbf{x}/C_k)$

$x \in \{v_1, v_2 \ldots v_M\}^D$

$x_1 \quad x_2 \quad \ldots \quad x_D$

$M \quad M \quad \quad M \longrightarrow [M^D - 1]$

the probabilities sum to 1

# Probabilistic Generative Models - Discrete

- The 'Naive Bayes' Assumption - feature values are treated as independent when conditioned on class $C_k$

$$p(\mathbf{x}/C_k) = \prod_{i=1}^{D} p(x_i/C_k) = \prod_{i=1}^{D} \left[ \pi_{ki}^{x_i} (1-\pi_{ki})^{1-x_i} \right]$$

$x_i$ is binary

$\Rightarrow$ either 0 or 1

$t_i = \left\{ \begin{matrix} c_1 \\ c_2 \end{matrix} \right\}$

$x_i = \{0, 1\}$

# Probabilistic Generative Models - Discrete

- Posterior probability

$$p(C_k/\mathbf{x}) = \frac{P(x/C_k)\,P(C_k)}{P(x)} = \frac{e^{a_k}}{\sum_{j=1}^{K} e^{a_j}}$$

$$a_k(\mathbf{x}) = \ln\left[P(x/C_k)\cdot P(C_k)\right]$$

$$= \ln\left[\prod_{i=1}^{D} \pi_{ki}^{x_i}(1-\pi_{ki})^{1-x_i} \cdot P(C_k)\right]$$

$$= \underbrace{\sum_{i=1}^{D}\left(x_i \ln \pi_{ki} + (1-x_i)\ln(1-\pi_{ki})\right]}_{\text{linear in } x} + \ln P(C_k)$$

# Probabilistic Generative Models - Discrete

- Analogous results can be obtained for non-binary components
    - Exercise!
- Derive the ML estimates for the Binary case
    - Exercise!

assume independence conditioned on the class.

Also, utilize 1-of-L encoding for the values of the components.

Next
Discriminative Models