GENERATIVE VS DISCRIMINANT CLASSIFIERS

TRAINING DISCRIMINATIVE : $\sum_N \log p(y_i | x_i, \theta)$ CONDITIONAL LOG LIKELIHOOD

TRAINING GENERATIVE: $\sum \log p(y_i, x_i | \theta)$ JOINT LOG LIKELIHOOD

IN GENERAL DISCRIMINATIVES ARE MORE ACCURATE BECAUSE 'EASIER JOB'

EASIER TO FIT : GENERATIVE

ADD CLASSES: GENERATIVE B/C CLASS CONDITIONAL ARE INDEPENDENT

MISSING FEATURES: GENERATIVE

SEMI-SUPERVISED : GENERATIVE

OUTPUT → INPUT : GENERATIVE

FEATURE PREPROCESSING: DISCRIMINATIVE, IE BASIS FUNCTION EXPANSION
GEN. NOT SO WELL B/C CORRELATIONS

PROBABILITY CALIBRATION : DISCRIMINATIVE
GEN: XTREME POSTERIORS | STRONG ASSUMPTIONS

MISSING DATA

SOME INPUTS OF (x) ARE NOT OBSERVED : $R_i \in \{0, 1\}$ DATA/NO DATA $p(x_i, R_i | \theta \phi) = p(R_i | x_i, \phi) p(x_i | \theta)$

MCAR, MAR, NMAR

DATA ONLY MISSING IN ~~TRAI~~ TEST SET: MARGINALIZE OUT THE MISSING FEATURES $p(y=c | x_{2:D}, \theta) \propto p(y=c|\theta) \sum p(x_1, x_{2:D} | y=c, \theta)$

IF NAIVE BAYES $\sum p(x_1, x_{2:D} | y=c, \theta) = \prod^{D} p(x_j | \theta_{jc})$ → WE SIMPLY IGNORE MISSING FEATURES

DATA ALSO MISSING IN TRAINING SET: HARDER, IE EM ALGORITHM

SEMI-SUPERVISED LEARNING: LABELS ALSO MISSING SOMETIMES AT TRAINING TIME