Intuition

How about if we inferred properties of a subset of the training inputs…

...but not of the whole class?

In a nutshell: given a gender classifier, infer race of people in Bob's photos

6

In a nutshell: given a gender classifier, infer race of

training inputs…

...but not of the whole class?

How about if we inferred properties of a subset of the

people in Bob’s photos

7