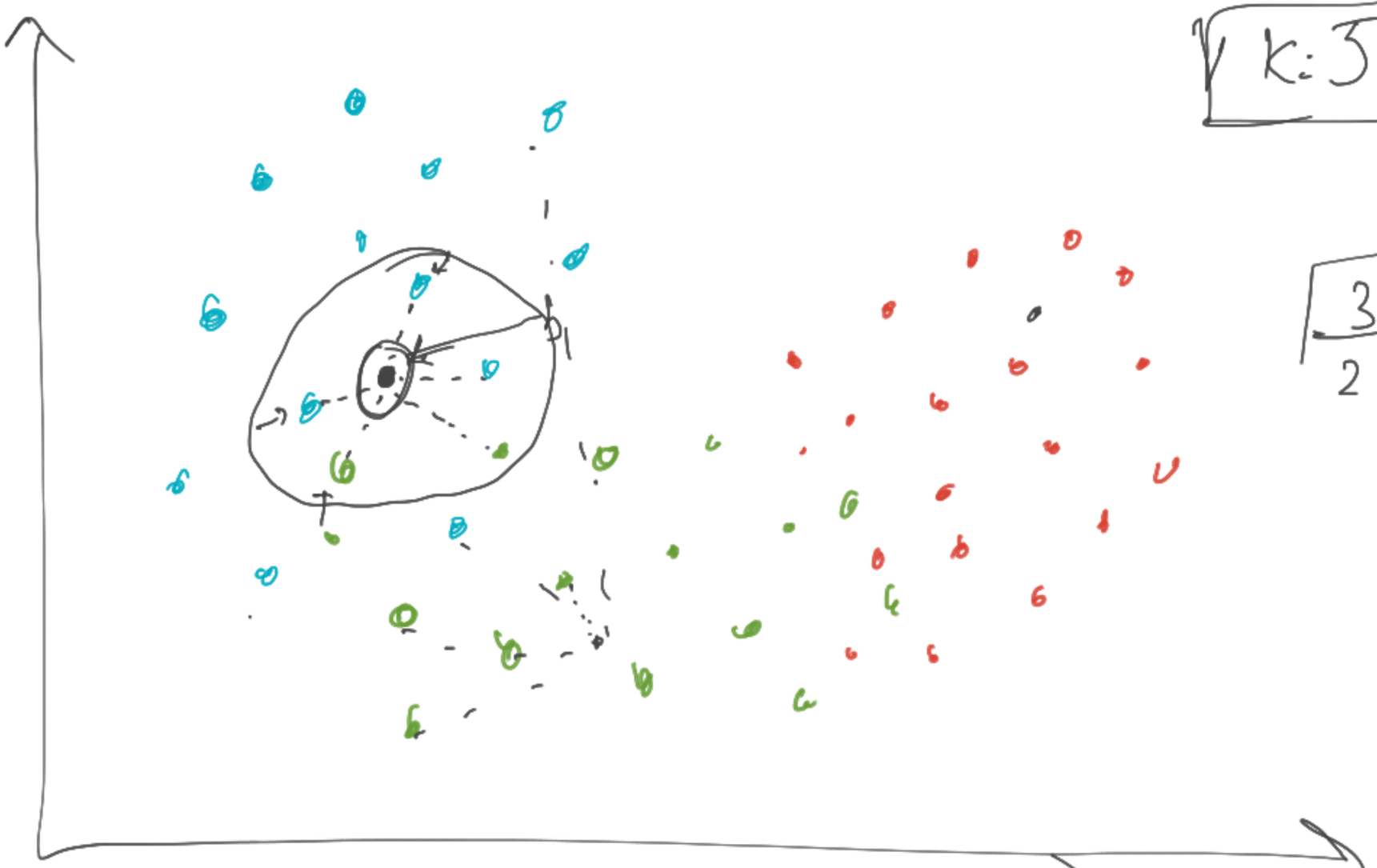
KNN → Regression
KNN → classification

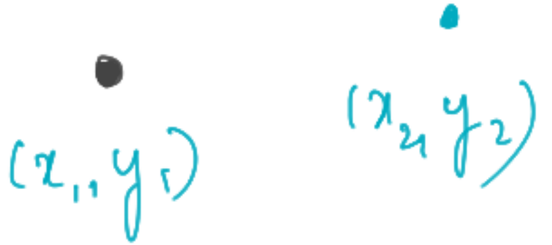
1) test point
2) find nearest pt
3) Note on the label 7
   mean of the neig.




non parametric

Euclidian dist: $p = 2$

$$= \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$




Manhattan dist $p = 1$

$$|x_2 - x_1| + |y_2 - y_1|$$

3, 5, 7, 11, .. 99 except

1000

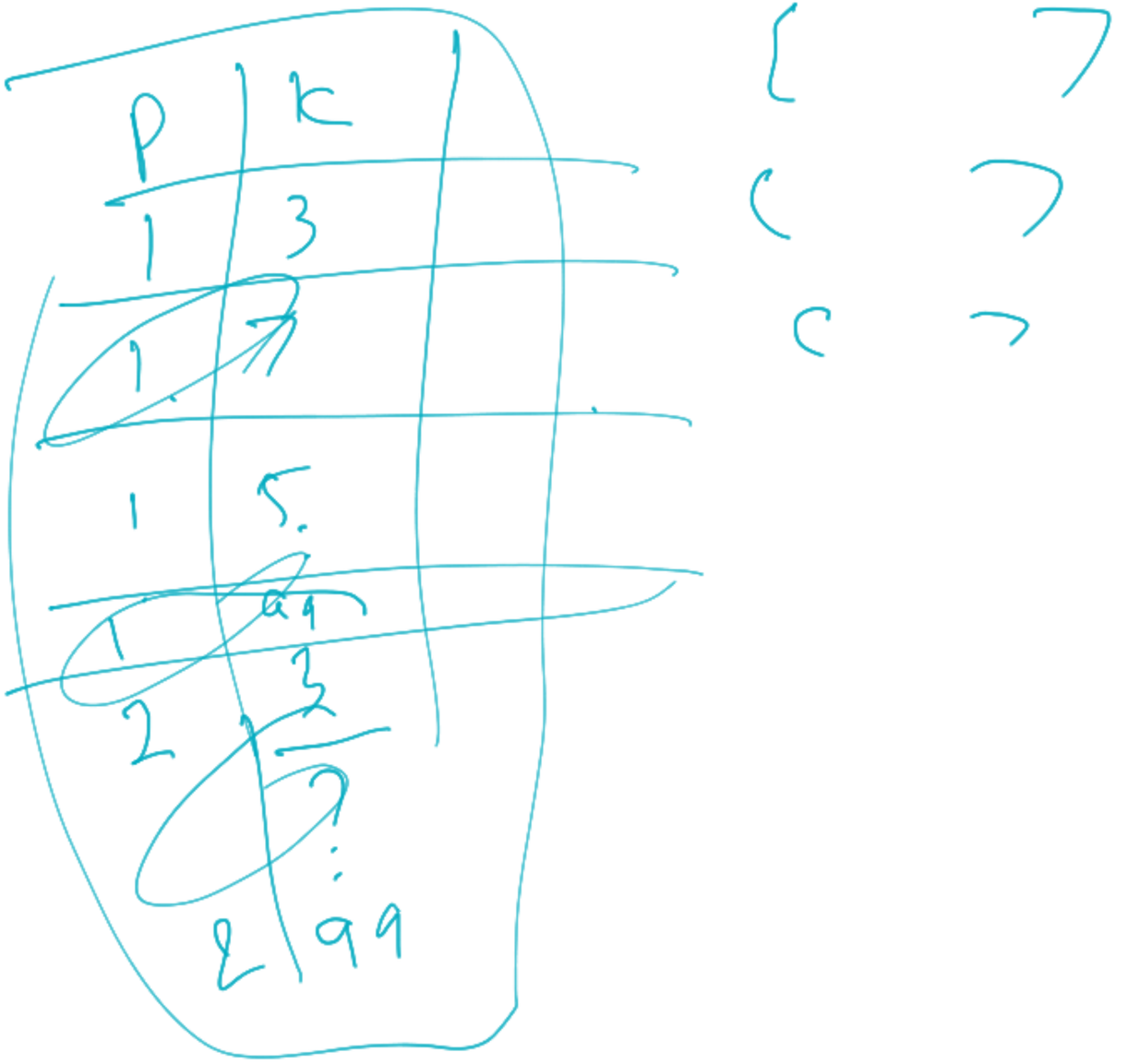
Cross validation

k: 3 .. 99

p: 1 .. 3

- ✓ GridSearchCV
- ✓ RandomSearchCV

best param



Adv:

1) Non parametric
2) no training (just testing phase)
3) easy > both r/c
4) easy for multi-class classification
5) only hyperparameter is k & p

Dis:

1) Lazy learner (test stage is slower)
2) Sensitive to outliers
3) Impacted by unbalanced dataset
4) Low accuracy in high dimensions.
5) Feature scaling required

| 1 | 2 | 3 |
|---|---|---|
| age | salary | grad % |
| 18-60 | 5-5, thousands/ lakhs | 35-100% |