Homework #7

equal
x > y

k-NN	0.8	0.7	0.8	0.7	0.8
Logistic	0.9	0.7	0.8	0.9	0.7

Given 5 variants of the training set obtained by bootstrap, the accuracies of k-NN and Logistic Regression are given above.

Compute the two mean accuracies and the Vargha-Delaney effect size. Based on the two means and the effect size, knowing that *p*-value = 0.6 and power = 0.33, what is your conclusion on the superiority of k-NN vs Logistic (or Logistic vs k-NN)?

Since the accuracies were obtained through by bootstrapping then this is the case of paired data.

K-NN mean

$$\frac{3\times 0.8 + 2\times 0.7}{5} = 0.76$$

Logistic mean

$$\frac{2\times 0.9 + 2\times 0.7 + 0.8}{5} = 0.8$$

$\hat{A}_{12} = P(x>y)$ so we have to see pair-wise how many are equal and how many $x>y$

equal = 2
$(x>y) = 1$

$$\hat{A}_{12} = \frac{1+0.5\times 2}{5} = \frac{2}{5}$$

$$2\left|\hat{A}_{12} - 0.5\right| = 2\left|\frac{2}{5} - \frac{1}{2}\right| = 2\left|\frac{4-5}{10}\right| = 0.2$$

→ Small d-effect size

The p-value is 0.6>0.05 so we cannot reject the null hypothesis. Since the power is 0.33 < 0.8 cannot accept it either because we need more data. So the result is inconclusive

Homework #8

	I1	I2	I3	I4	I5
x1	1	5	3	4	1
x2	0.9	0.5	0.6	1.0	0.8
y	A	B	B	A	A

The training set X = {I1, I2, I3, I4, I5} contains input vectors with two attributes x1, x2.

Evaluate the quality of splitting of X using Gini Impurity for the two following candidate splitting conditions:

```
a1: x1 < 5
a2: x2 < 0.8
```

$$GI(X) = 1 - \sum_{c \in C} p_c^2 \quad , \quad p_c = \frac{|\{x \in X \wedge x \in C|}{|X|}$$

$$GI(X, a) = p_a GI(X_a) + (1 - p_a) GI(X_{\tilde{a}}), \quad p_a = \frac{|\{x \in X \wedge a(x)\}|}{|X|}$$

$a(x) :=$ x meets condition a

$\underline{\underline{a_1}}$: $x_1 < 5$

$$a_1 = \{I_1, I_3, I_4, I_5\}, \quad \tilde{a}_1 = \{I_2\}$$

$$p_{a_1} = \frac{4}{5}, \quad p_{\tilde{a}_1} = \frac{1}{5}$$

$$GI(X_a) = 1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2 = 1 - \frac{10}{16} = \frac{3}{8}$$

$$GI(X_{\tilde{a}}) = 1 - \frac{0}{1}^2 - \left(\frac{1}{1}\right)^2 = 0$$

$$GI(X, a_1) = \frac{4}{5} \times \frac{3}{8} + \frac{1}{5} \times 0 = \frac{3}{10}$$

$\underline{\underline{a_2}}: x_2 < 0.8$

$a_2 = \{ l_2, l_3 \}$, $\tilde{a}_2 = \{ l_1, l_4, l_5 \}$

$p_{a_2} = \frac{2}{5}$, $p_{\tilde{a}_2} = \frac{3}{5}$

$GI(X_{a_2}) = 1 - 0^2 - \left(\frac{2}{2}\right)^2 = 0$

$GI(X_{\tilde{a}_2}) = 1 - \left(\frac{3}{3}\right)^2 - 0^2 = 0$

$GI(X, a_2) = \frac{2}{5} \times 0 + \frac{3}{5} \times 0 = 0$

Homework #9

	v1	v2	v3	v4	v5
x1	1	4	2	3	1
x2	0	0	1	1	1
y	A	B	B	A	A

Train a naive Bayes classifier using the 5 training vectors v1, …, v5 (whose 2 dimensions are attributes x1 and x2, both categorical) with Laplace smoothing and use it to classify:

```
v6 = <4, 1>
```

$\frac{\alpha}{\beta}$ $\begin{cases} \alpha := \text{number of times of attribute given class } +1 \\ \beta := \text{number of times class appears + number of different attributes} \end{cases}$

X_1	1	2	3	4
A	$\frac{2+1}{3+4}$	$\frac{0+1}{3+4}$	$\frac{1+1}{3+4}$	$\frac{0+1}{3+4}$
B	$\frac{0+1}{2+4}$	$\frac{1+1}{2+4}$	$\frac{0+1}{2+4}$	$\frac{1+1}{2+4}$

X_2	0	1
A	$\frac{1+1}{2+3}$	$\frac{2+1}{2+3}$
B	$\frac{1+1}{2+2}$	$\frac{1+1}{2+2}$

$P(A) = \frac{3}{5}$

$P(B) = \frac{2}{5}$

$V_6 = \langle 4, 1 \rangle$

$$P(A \mid V_6) = P(X_1 = 4 \mid A)\, P(X_2 = 1 \mid A)\, P(A)$$
$$= \frac{1}{7} \cdot \frac{3}{5} \cdot \frac{3}{5} = \frac{9}{175} \approx 0.051$$

$$P(B \mid V_6) = P(X_1 = 4 \mid B)\, P(X_2 = 1 \mid B)\, P(B)$$
$$= \frac{2}{6} \cdot \frac{2}{4} \cdot \frac{2}{5} = \frac{1}{15} = 0.2$$

Then V_6 is classified as B.

Homework #10

After training, an SVM classifier has the following 4 support vectors:

```
x1 = <1, 3>, y1 = +1
x2 = <2, 5>, y2 = +1
x3 = <1, 1>, y3 = -1
x4 = <2, 3>, y4 = -1
```

Assuming the Lagrange coefficients c_i to be all +1 and b = 0, how does a linear SVM classifies the following vectors?

```
x5 = <0, 1>
x6 = <1, -1>
```

Keeping the same support vectors and Lagrange coefficients, and b = 0, does the classification change when using a radial basis kernel with σ = 1?

$C :=$ SVM classifier

$$C(x) = \operatorname{sgn}\left(\sum_i y_i c_i (x_i \cdot x) - b\right)$$

dot products

	x_1	x_2	x_3	x_4
x_5	3	5	1	3
x_6	-2	-3	0	-1

$$C(x_5) = \operatorname{sgn}(1 \cdot 1 \cdot 3 + 1 \cdot 1 \cdot 5 - 1 \cdot 1 \cdot 1 - 1 \cdot 1 \cdot 3 - 0) = \operatorname{sgn}(4) = +$$

$$C(x_6) = \operatorname{sgn}(1 \cdot 1 \cdot (-2) + 1 \cdot 1 \cdot (-3) - 1 \cdot 1 \cdot 0 - 1 \cdot 1 (-1) - 0) = \operatorname{sgn}(-4) = -$$

$$K(x_i, x) = \exp\left(-\frac{d(x_i, x)^2}{2\sigma^2}\right)\Big|_{\sigma=1} = \exp\left(-\frac{d(x_i, x)^2}{2}\right)$$

C_K: Non-linear SVM with radial basis kernel and $\sigma = 1$

$$C_K(x) = \operatorname{sgn}\left(\sum_i y_i c_i K(x_i, x) - b\right)$$

Homework #11

Given the training set:

```
x1 = <1, 3>
x2 = <2, 4>
x3 = <1, 1>
x4 = <2, 3>
```

1 2

The Lagrange coefficients and the intercept of a one-SVM classifier are c = [0.0004, 0, 0, 0], ρ = 0.0002. Determine if `x5 = <0, 1>` and `x6 = <1, 0>` are outliers, assuming a Gaussian kernel with σ = 1.

If we use agglomerative clustering with complete linkage, cutting the hierarchy at k = 3, how are `x5` and `x6` classified? Are they outside or inside the closest cluster, based on the cluster's radius?

$$C_K(x) = \text{sgn}\left(\sum_i c_i K(x_i, x) - \rho\right)$$

$$K(x_i, x) = \exp\left(-\frac{d(x_i, x)^2}{2\sigma^2}\right)$$

$$C_K(x_5) = \text{sgn}(-0.0002) = -$$

$$C_K(x_6) = \text{sgn}(-0.0002) = -$$

The points are classified as anomalies.

<u>agglomerative w/ comple linkage</u>

distances^2

	x_1	x_2	x_3	x_4
x_1	0	2	4	1
x_2		0	10	1
x_3			0	5
x_5				0

$\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}$

$\{x_1, x_4\}, \{x_2\}, \{x_3\}$ or $\{x_1\}, \{x_2, x_4\}, \{x_3\}$

centroids				centroids	
$(\frac{3}{2}, 3)$ C_1	x_2	x_3	x_1	$(2, \frac{7}{2})$ C_2	x_3

Identifying anomalies w/ clusters

1. Anomalies don't belong to any cluster

$$d(x, c_i) > \rho(C_i), \quad \rho(C_i) = \max_{x \in C_i} d(x, c_i)$$

2. Anomalies is far from centrois c_i

$$d(x, c_i) > \alpha \rho(C_i)$$

3. Anomalies belong to small/sparse centroids

cluster C_i closest to x is small if $|C_i| < \beta N$

cluster C_i closest to x is sparse if $\rho(C_i) > \gamma \bar{\rho}$

γ → cluster radius; $\bar{\rho}$ → Average radius

distances

	C_1	x_2	x_3
x_5	$4+\frac{9}{4}$	13	1
x_6	$9+\frac{1}{4}$	17	1

x_5 and x_6 are assigned to x_3. The radius of x_3 is 0. Hence they are classified as anomalies by criteria 1.

There is no homework #12!!

Homework #13

Compute association rules from the following database of transactions, with min confidence and support threshold set to 0.75:

```
p q r
p s
p s
p s r
```

Determine the lift of the resulting association rules. Is any association rule interesting?

① first Consider only singletons with support ≥ 0.75

$$\left.\begin{array}{l} supp(\{p\}) = \frac{4}{4} = 1 \\ supp(\{q\}) = \frac{1}{4} \\ supp(\{r\}) = \frac{2}{4} = \frac{1}{2} \\ supp(\{s\}) = \frac{3}{4} \end{array}\right\} L_1 = \{\{p\}, \{s\}\}$$

② Combine singletons and assess support

$$Supp(\{p, s\}) = \frac{3}{4} \quad \} \quad L_2 = \{\{p, s\}\}$$

(3) For each of the sets L_i remove an item and make it a consequent, the compute confidence

Rule	Confidence
$\{\} \Rightarrow \{p\}$	$\frac{4}{4} = 1$
$\{\} \Rightarrow \{s\}$	$\frac{3}{4}$
$\{p\} \Rightarrow \{s\}$	$\frac{3}{4}$
$\{s\} \Rightarrow \{p\}$	$\frac{3}{3} = 1$

(4) See if we can merge roles with same antecedent

	Confidence	
$\{\} \Rightarrow \{p, s\}$	$\frac{3}{4}$	✓

Final set of roles:

$\{\} \Rightarrow \{p, s\}$

$\{p\} \Rightarrow \{s\}$

$\{s\} \Rightarrow \{p\}$

5) Given $A \Rightarrow B$, Lift is computed by $\dfrac{Conf(A \Rightarrow B)}{Supp(B)}$

Rule	conf	Supp(B)	Lift
$\{\} \Rightarrow \{p, s\}$	$\frac{2}{4}$	$\frac{3}{4}$	1
$\{p\} \Rightarrow \{s\}$	$\frac{3}{4}$	$\frac{3}{4}$	1
$\{s\} \Rightarrow \{p\}$	1	1	1

All have the same Lift, hence are equally interesting.