(a) Scored data set

87 85 82 79 74 73
68 66 61 39 28 12

(b) How $B(r)$ is calculated

Real cats

87 85 82
79 68 66

Labelled cats (r=67)

87 85 82
79 74 73
68

Choose (with replacement) n real cats and n labelled cats at random (discard if same image chosen for comparison) and compare scores
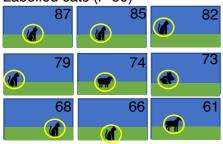
$$\underbrace{\begin{matrix}68\\>\\73\end{matrix}}_{0} > \underbrace{\begin{matrix}87\\>\\68\end{matrix}}_{1} > \underbrace{\begin{matrix}79\\>\\73\end{matrix}}_{1} \cdots > \underbrace{\begin{matrix}85\\>\\87\end{matrix}}_{0} > \underbrace{\begin{matrix}79\\>\\82\end{matrix}}_{0}$$

$B(r)$ is proportion of times real score score is more than labelled, i.e.
$$B(r) = (0 + 1 + 1 + \cdots + 0 + 0)/n$$

In this example: $B(65) \approx 0.49$
and $C(67) = 5/7 \approx 0.71$
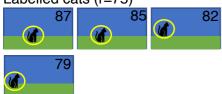
(c) Effect of changing threshold (r)

Labelled cats (r=60)

87 85 82
79 74 73
68 66 61

Decreasing r, increases the number of labelled cats and gives a larger value of $B(r)$. Real cats are unaffected by the choice of r.

In this example: $B(60) \approx 0.60$ and $C(60) = 6/9 \approx 0.66$

Labelled cats (r=75)

87 85 82
79

Increasing r, decreases the number of labelled cats and gives a smaller value of $B(r)$.

In this example: $B(75) \approx 0.49$
and $C(75) = 4/4 = 1$