Quantifying Classification Performance

In 1990, Wolberg and Mangasarian (1990) developed a technique to accurately diagnose breast masses using only visual characteristics of the cells within the tumor. A sample is placed on a slide, and characteristics of the cellular nuclei within the tumor are examined under a microscope to determine whether the cancer cells are *benign* or *malignant*.

Suppose oncologists use our fitted additive regression model to help them make this determination on the basis of the average radius of the cell nuclei and their concavity. They use a classification cut-off of $\pi_0 = 0.5$; when applied to the 569 individuals in the sample, this produces the following classification table:

	Classified as Malignant	Classified as Benign	Total
Malignant	188	24	212
Benign	22	335	357

1. What is the **sensitivity** corresponding to this choice of cut-off?

2. What is the **specificity** corresponding to this choice of cut-off?

3. Another measure of predictive performance is the **correct classification rate** or **accuracy**. This measures the overall proportion of individuals who were correctly classified (across both tumor types). What is the accuracy of our classifier?

4. Accuracy may feel like a more intuitive measure of predictive performance than the sensitivity or the specificity. However, can you think of a scenario in which reporting the accuracy alone might be misleading?

Classification Performance Under Different Cut-Offs

How would our logistic regression model's performance differ with different choices for the classification cut-off (π_0)?

1. Suppose you decide to set $\pi_0 = 0$, so that all tumors with predicted probabilities $\hat{\pi}_i > \pi_0 = 0$ are classified as malignant. What does the corresponding classification table look like? What are the sensitivity and specificity?

	Classified as Malignant	Classified as Benign	Total
Malignant			212
Benign			357

2. How do your answers change if you take $\pi_0 = 1$ instead?

	Classified as Malignant	Classified as Benign	Total
Malignant			212
Benign			357

3. In general, how do you think the sensitivity changes as we increase the value of π_0? What about the specificity?

4. What about the false alarm (also known as false positive) rate (a.k.a., $1 - specificity$)?

References

Wolberg, William H, and Olvi L Mangasarian. 1990. "Multisurface Method of Pattern Separation for Medical Diagnosis Applied to Breast Cytology." *Proceedings of the National Academy of Sciences* 87 (23): 9193–96. https://doi.org/10.1073/pnas.87.23.9193.