

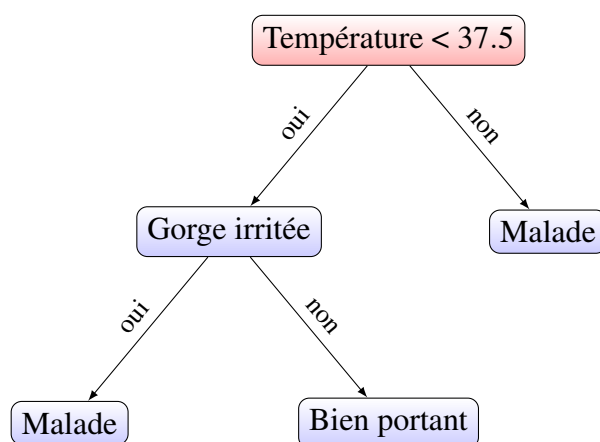
## TD n° 1

### Arbre de décision

**Exercice 1** On dispose d'un échantillon de 200 patients. On sait que 100 sont malades ( $m$ ), et les 100 autres sont bien portants ( $bp$ ). On dispose en outre des informations suivantes :

	gorge irritée		gorge non irritée	
température < 37.5	6 bp	37 m	91 bp	1 m
température ≥ 37.5	2 bp	21 m	1 bp	41 m

Soit l'arbre de décision de la figure suivante :



1. Calculer, pour l'arbre de décision donné, le gain d'information à chaque nœud en utilisant la fonction d'entropie.
2. Considérons l'arbre vide. Nous avons le choix entre choisir l'attribut "température < 37.5" et l'attribut "gorge irritée". Lequel doit-on choisir pour maximiser le gain ?

**Exercice 2** Soit un jeu de données de 9 patients contenu dans le tableau suivant. Les attributs sont "Fievre", "Douleur" et "Toux". La classe est "Maladie".

Fievre	Douleur	Toux	Maladie
oui	Abdomen	non	Appendicite
non	Abdomen	oui	Appendicite
oui	gorge	non	rhume
oui	gorge	oui	rhume
non	gorge	oui	mal de gorge
oui	non	non	aucune
oui	non	oui	rhume
non	non	oui	refroidissement
non	non	non	aucune

- Déterminez l'arbre de décision de ces exemples. Détaillerez à chaque fois le calcul de l'entropie et du gain.